An investigation into the genetic risk of age-related hearing impairment

Wells, Helena

Awarding institution:
King's College London

An investigation into the genetic risk of age-related hearing impairment

Helena R.R. Wells

A thesis presented for the degree of

Doctor of Philosophy

Department of Twin Research and Genetic Epidemiology
King's College London, United Kingdom
2020

Abstract

Age-related hearing impairment (ARHI) is the most common sensory impairment in the ageing population. Disabling hearing loss affects a third of people aged 65 and is a major risk factor for dementia. Little is currently known about genetic risk factors and the underlying pathology of ARHI, making it a challenge to develop accurate diagnoses and treatment strategies. ARHI is a thought to be a polygenic, heterogeneous trait and so under the common trait common variant hypothesis, large studies are required to achieve the necessary statistical power to identify multiple genetic variants with small effect sizes. Previous studies have been limited by reduced sample sizes and the absence of a comprehensive, consistent definition of ARHI.

In this thesis, the UK Biobank (UKBB) resource was used to perform two genome-wide association–studies (GWAS) that were a magnitude larger than previous ARHI GWAS (n>250,000) and various methods were used to validate and investigate the subsequent associations. The results support the use of GWAS to uncover putative ARHI genetic risk variants. Forty-four independent genome-wide significant loci ($p < 5E-08$) were identified from the two GWAS studies, a 9-fold increase in the number of loci previously associated with common adult hearing impairment. A quarter of the identified loci contain genes that are implicated in other forms of hearing loss, while the remaining are novel gene candidates. The work demonstrates the viability of using self-report measures of adult hearing impairment for genetic discovery in large samples for the first time. Further, genetic correlation analysis revealed strong positive correlations with multiple personality and psychological traits for the first time. Multiple associations were validated via a replication meta-analysis, and tissue-specific expression was observed with immunohistochemistry for three candidate genes at significantly associated loci. All associated loci were evaluated for use in functional analysis, providing a framework for future studies.

Acknowledgements

Firstly, I would like to thank Action on Hearing Loss and the family that generously donated funds for this project, it was a pleasure to meet you. I would also like to thank the many volunteers of UK Biobank, TwinsUK and ELSA, without whom none of the findings described in this thesis would have been possible.

Secondly to my colleagues at Kings, where I spent the first half of my project. Thank you to my supervisor Frances Williams, for giving me this opportunity, for your guidance, fruitful collaborations and for your valuable input to this thesis and multiple manuscripts. Thank you to Maxim Freidin for your direction throughout this project, for your contributions and for letting me believe that no question is a stupid question. I know how lucky your current and future students are. Thank you to my thesis panel; Claire Steves, Pirro Hysi and Lisa Nolan for your friendly support, valued advice and input. Lastly, thanks to the many colleagues turned friends, especially my fellow PhD students for navigating the last few years with me.

Thirdly I wish to thank my colleagues at UCL where I spent the second half of my project. Thanks to my supervisor Sally Dawson, for sharing your expertise, for your contributions and for keeping me on track with our many honest discussions. I felt like your door was always open (apologies if it wasn't!). Thank you to my colleagues at UCL for being so welcoming and for your academic guidance, especially Claudia and Fatin.

Finally, I would like to thank my friends and family for their encouragement, laughs and for maintaining my perspective. Thank you to my partner, Mark, I couldn't be more grateful for your support and for pushing me to reach this stage with humour and kindness. Here's to weekends sans thesis! To my parents and Miranda, I cannot thank you enough for your continuous support and unwavering belief in me. Thank you for my education and for fuelling my determination, this is as much your achievement as it is mine.

Publications

Publications contributed to during the writing of this thesis:

Wells, H. R. R., Freidin, M. B., Zainul Abidin, F. N., Payton, A., Dawes, P., Munro, K. J., Morton, C. C., Moore, D. R., Dawson, S. J. & Williams, F. M. K., GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. 3 Oct 2019, In : American Journal of Human Genetics. 105, 4, p. 788-802.

Wells HRR, Newman TA, Williams FMK. Genetics of age-related hearing loss. J Neurosci Res. 2020 Jan 27. doi: 10.1002/jnr.24549.

Dunbar, L. A., Patni, P., Aguilar, C., Mburu, P., Corns, L., **Wells, H. R. R.,** Delmaghani, S., Parker, A., Johnson, S., Williams, D., Esapa, C. T., Simon, M. M., Chessum, L., Newton, S., Dorning, J., Jeyarajan, P., Morse, S., Lelli, A., Codner, G. F., Peineau, T. & 12 others, Clarin-2 is essential for hearing by maintaining stereocilia integrity and function, 1 Sep 2019, In : EMBO Molecular Medicine. 11, 9, e10288.

Nagtegaal, A. P., Broer, L., Zilhao, N. R., Jakobsdottir, J., Bishop, C. E., Brumat, M., Christiansen, M. W., Cocca, M., Gao, Y., Heard-Costa, N. L., Evans, D. S., Pankratz, N., Pratt, S. R., Price, T. R., Spankovich, C., Stimson, M. R., Valle, K., Vuckovic, D., **Wells, H.**, Eiriksdottir, G. & 22 others, Genome-wide association meta-analysis identifies five novel loci for age-related hearing impairment., 1 Dec 2019, In : Scientific Reports. 9, 1, 15192.

Cherny, S. S., Livshits, G., **Wells, H. R. R.**, Freidin, M. B., Malkin, I., Dawson, S. J. & Williams, F. M. K., Self-reported hearing loss questions provide a good measure for genetic studies: a polygenic risk score analysis from UK Biobank, 20 Mar 2020, In : European journal of human genetics : EJHG.

Contents

List of Figures

List of Tables

Abbreviations commonly used throughout the thesis

ARHI	Age-related hearing impairment
ANF	Auditory nerve fibre
dB	Decibels
DTT	Digit triplet test
ELSA	English longitudinal study of ageing
HAid	Hearing aid phenotype used in association study, chapter 4
HC	Hair cell
HDiff	Hearing difficulty phenotype used in association study, chapter 4
HL	Hearing loss
GWAS	Genome-wide association study
IHC	Inner hair cell
LD	Linkage disequilibrium
MET	Mechanoelectrical transducer
NSHL	Non-syndromic hearing loss
NIHL	Noise-induced hearing loss
OHC	Outer hair cell
PC	Principal component
PTA	Pure tone audiometry
QC	Quality control
SGN	Spiral ganglion neuron
SHL	Syndromic hearing loss
SIN	Speech in noise
SNP	Single nucleotide polymorphism
SNR	Speech to noise ratio
SRT	Speech reception threshold
UKBB	UK Biobank

Chapter 1 - Introduction

1.1 Auditory function

The auditory system spans multiple organs, tissues and cell types between and including the outer ear and the auditory cortex as depicted in Figure 1.1. The outer, middle and inner ear make up the peripheral auditory system, which is largely encased in the temporal bone; a hard bone that constitutes part of the side and base of the skull. The central auditory system comprises of structures from the cochlear nucleus to the primary auditory cortex and is largely located in the brain. In a fully functioning auditory system, the peripheral and central systems work independently and in conjunction with one another. During sound processing, the afferent system (Figure 1.1,(b)) runs in the direction from the outer ear to the primary auditory cortex, while the efferent system runs from the primary auditory cortex to the cochlea along a similar route.

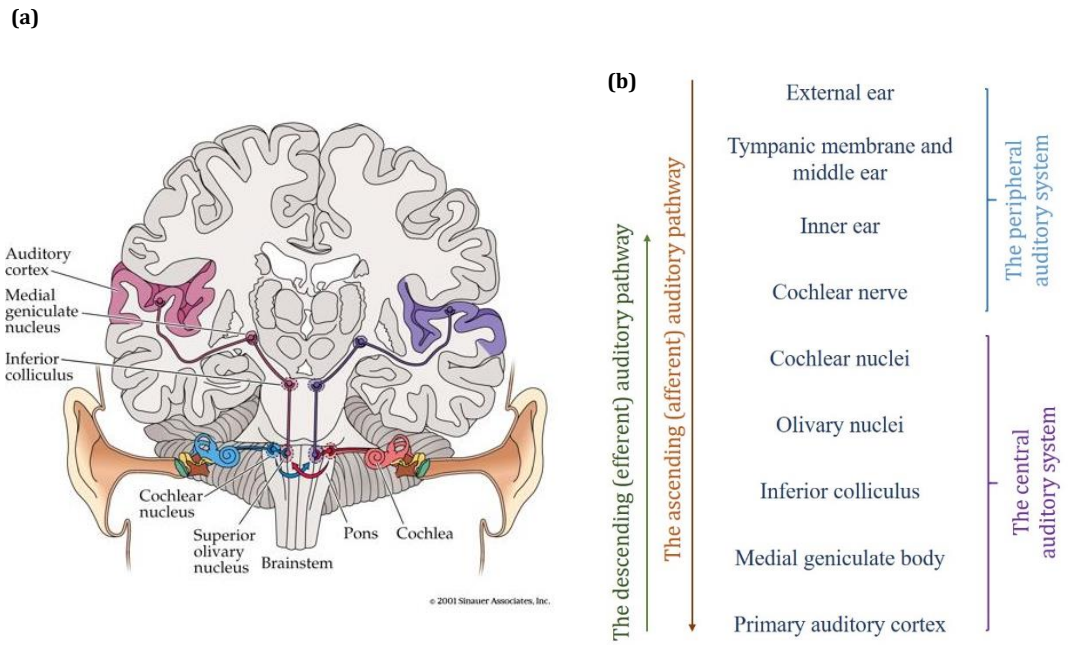

Figure 1.1. (a) Ascending auditory pathway. (b) Key structures of the auditory pathway.

(a)This diagram illustrates the path of the ascending auditory pathway. All auditory structures are coloured. Those which are not labelled can be identified by these colours on the left-hand side of the diagram: beige, outer ear; green yellow and brown, middle ear; blue, inner ear. Figure from Purves, Dale et al., 2001[1].
(b)The structures are presented in the order of the afferent pathway. The peripheral and central systems and ascending and descending pathways are discussed in detail later in the introduction. The colours do not correspond to the colours in Figure 1.1.

The peripheral auditory system: The outer and middle ear

The peripheral auditory system consists of the outer and middle ear, the cochlea and the auditory nerve. Three separate organs make up the ear itself; the outer ear, middle ear and inner ear, which work in combination to enable balance (vestibular system) and noise perception (peripheral auditory system). The outer ear comprises of the pinna (visible ear), the external auditory meatus (the ear canal) and the lateral portion of the tympanic membrane (the eardrum)[2], as displayed in Figure 1.2. The outer ear functions to channel acoustic stimuli toward the middle ear, via specialised crests and depressions and through the external auditory meatus (Figure 1.2). Skin covers the cartilage and the bone of the outer ear, and is partially covered with hair follicles and a wax that is produced by the sebaceous glands[2].

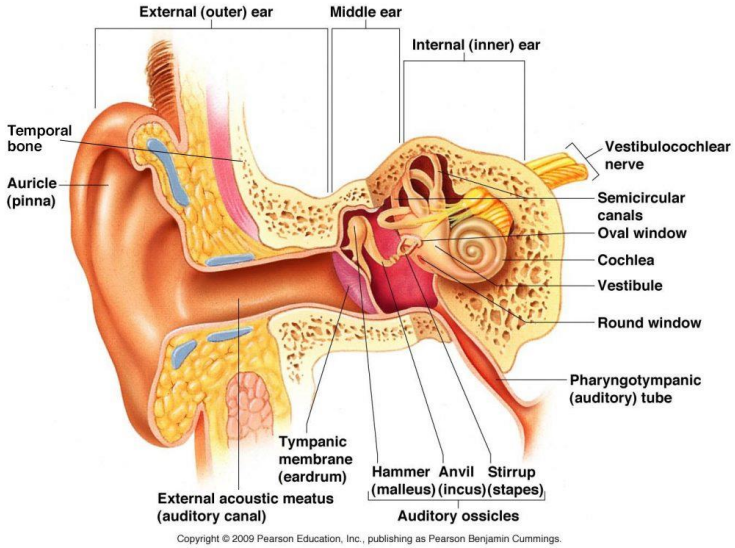

Figure 1.2. Anatomy summary of the outer, middle and inner ear.
https://www.pacificneuroscienceinstitute.org/eye-ear/hearing/overview/

The outer ear funnels the sound stimuli toward the tympanic membrane (ear drum), where it is converted into kinetic energy. The middle ear lies behind this tympanic membrane and contains three bones, the malleus, incus and stapes, that are collectively known as the ossicles. The medial wall of the inner ear comprises of the temporal bone and two membrane-covered openings; the oval window and the round window (Figure 1.2). When the sound stimuli reach the tympanic membrane, both the tympanic membrane and in turn the ossicles, vibrate precisely in response to the frequency and intensity of the received acoustic signal. The ossicles here work to amplify the acoustic signal received by the middle ear as the energy reaches the interface between the middle

and inner ears. Here the stapes connects to the oval window, the membrane between the air-filled middle ear and the fluid-filled inner ear.

Malformations of the outer ear may be inherited, result from a spontaneous genetic alternation or be acquired from environmental exposures *in utero* or later in life such as physical trauma, and are collectively termed 'conductive' hearing loss. For example, microtia is a condition that occurs when defects arise during the development of the pinna affecting its size, orientation, position and/or relief pattern[3]. These defects are graded on a scale from I to III based on the severity of the malformation and are generally treated with corrective surgeries. Structural malformations can also occur at the external auditory canal, rendering it narrowed or partially developed. These cases are graded on a scale from A-C based on the extent of the obstruction[3]. Conductive hearing loss can also be caused by malformations that affect the conformation of structures within the middle ear, such as deformations or fixation of the ossicles (including otosclerosis[4]), a narrowing of the tympanic cavity and, or the conformation of the round or oval windows. Mutations in the *POU3F4* gene transcript for example, result in a form of X-linked HL, causing morphological malformations of the temporal bone including either the internal auditory canal, vestibular aqueduct, modiolus and, or vestibule[5].

The peripheral auditory system: The inner ear

The inner ear (or the 'labyrinth') contains the cochlea and the peripheral vestibular system. The vestibular system functions to maintain balance and spatial orientation[6]. The cochlea receives and relays acoustic stimuli from the middle ear to the auditory nerve. The membranous oval window acts as a barrier between the air-filled middle ear and fluid-filled cochlea of the inner ear. As the membrane of the oval window vibrates, the fluid of the cochlea takes on the amplified energy of the acoustic stimuli from the middle ear.

The cochlea is a spiral structure that holds the organ of Corti, as depicted in Figure 1.3. The organ of Corti is the site of mechanotransduction; where kinetic energy is transformed to electrochemical activity via specialised mechanoreceptor sensory hair cells (HCs) [7]. The resulting stimuli is received by higher auditory structures via spiral ganglion neurons and the auditory nerve (Figure 1.3, A). Dysregulation of or damage to the structures of the inner ear therefore leads to forms of 'sensorineural' hearing loss such as ARHI.

The cochlea consists of three fluid filled chambers; the Scala Tympani, Scala Vestibuli and Scala Media. The Scala Tympani and Scala Vestibuli contain perilymph (0 mV, low K^+

concentration) and the middle chamber, the Scala Media, contains endolymph (80 mV, the endocochlear potential, high K^+ concentration) [7]. Figure 1.3, A and Figure 1.4 illustrate the configuration of these three chambers. The ionic potential of these fluids is termed the 'endocochlear potential' and is predominantly generated and maintained by the stria vascularis. The stria vascularis is a structure that lines part of the Scala Media chamber, Figure 1.4, and recycles K^+ ions in the endolymph via strial marginal cells[8]. Maintenance of this potential is critical for successful mechanotransduction via mechanoelectrical transduction (MET) channels near the tips of HC stereocilia (Figure 1.5).

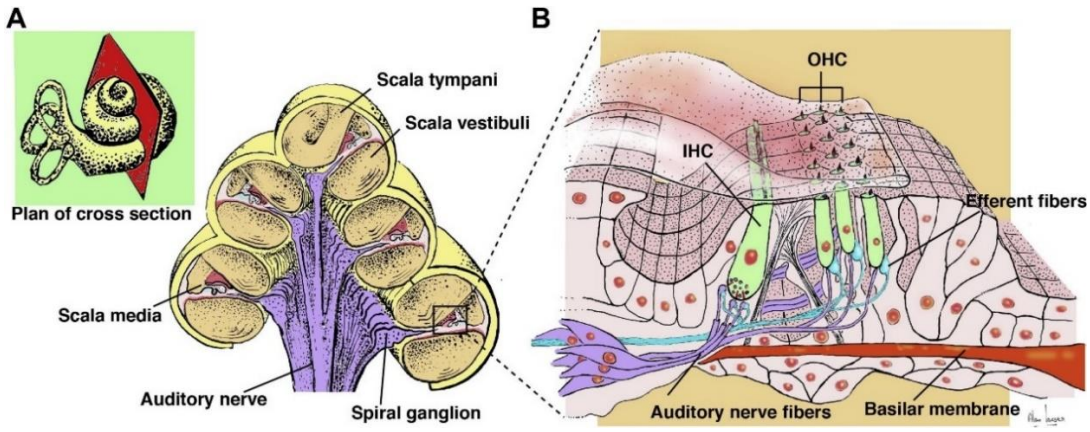

Figure 1.3. Structure of the cochlea and the organ of Corti.

(A) Cross-section of the cochlea (top inset shows diagram of the whole cochlea and sectioning scheme). The three compartments composing the cochlea are indicated (scala tympani, scala vestibuli and scala media). The organ of Corti is located in the scala media. The spiral ganglion comprises the somas of auditory nerve neurons innervating hair cells. (B) Structure of the organ of Corti. Inner (1) and outer (2) hair cells are indicated. Afferent dendrites belonging to auditory nerve neurons are indicated in purple, and medial olivary complex neurons in light blue. Some efferent fibers (light blue) also make axo-dendritic contacts with afferent boutons. Some afferent fibers represent a small proportion (type II) of afferent contacts on OHC. Image by Alan Larsen. Figure composition and legend from Goutman et al., 2015.[7].

The stria vascularis maintains the endolymph via a network of ion channels, gap junctions and tight junctions, which either permit or restrict the flow of ions into and out of the scala media. Key gap junction proteins are encoded by *GBJ2* and *GBJ6*, and aid communication between the intermediate and basal cells of the stria vascularis via ion transport. Mutations in *GJB2* are the most common cause of non-syndromic hearing loss in a number of populations[9,10], as discussed in section 1.2. The passive movement of ions between the marginal and basal cell layers of the stria vascularis is prohibited by tight junctions such as those encoded by *CLDN14* and *MARVELD2.* These tight junctions can also be a site of pathology; forms of autosomal recessive non-syndromic hearing loss are caused by mutations within these two genes[11,12].

The HCs are the sensory cells of the organ of Corti. The inner hair cells (IHCs) relay the primary stimuli via mechanotransduction. The outer hair cells (OHCs) primarily amplify and fine-tune the stimuli by electromechanical feedback[13] and electromotility, a unique property of OHCs. IHCs spiral around the structure in single a row of ~3,500 cells while the OHCs spiral around the structure in three rows of ~12,00 cells. 'Supporting' cells surround the sensory cells in the organ of Corti. Phalangeal cells support the IHCs and the Deiters' cells support the OHCs, along with supporting cells such as the Hensen and Claudian cells (Figure 1.4). Between the rows of IHCs and OHCs lie the Pillar cells, that form a structure termed the 'tunnel of Corti'.

Sensory HCs are organised tonotopically along the length of the cochlea; higher frequencies are transduced at the basal end of the cochlea and lower frequencies at the apical end of the cochlea. The basilar membrane, upon which the organ of Corti sits (Figure 1.3, B &, Figure 1.4), is specialised in function such that it vibrates according to specific fluid movement, resulting in frequency tuning. The membrane is wider and more flexible toward the apical end of the cochlea that responds to lower frequency acoustic stimuli, and narrower and stiffer at the basal end of the cochlea that responds specifically to high frequency acoustic stimuli[14].

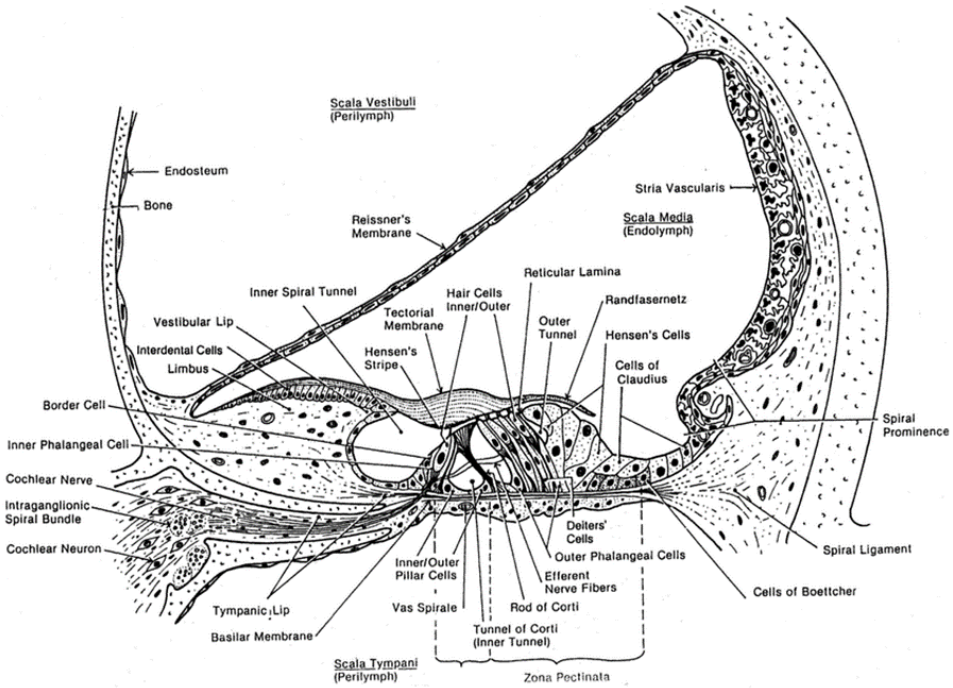

Figure 1.4. Schematic of the organ of Corti. Davis and Silverman, 1970[15].

At the apical end of the sensory HCs, lie 20-500 mechanosensing stereocilia[13] that are anchored to the apex of the HC by a structure called the cuticular plate. Actin filaments at the base of the stereocilia form rootlets in the cuticular plate, organised into dense

bundles by TRIO and F-actin binding protein, encoded by *TRIOBP*. Deleterious mutations in *TRIOBP* can therefore prevent the correct organisation of these anchoring bundles, resulting in a form of non-syndromic recessive HL[16].

The stereocilia are physically connected to adjacent stereocilia via cross links, made up of proteins such as espin, which provide strength and rigidity to the links. Deleterious mutations in the *ESPN* transcript, that codes for espin, are implicated in multiple types of non-syndromic hearing loss and can present in conjunction with balance problems, indicative of vestibular dysfunction[11]. The stereocilia are also connected via tip links (Figure 1.5A) which create synchronised movement of the stereocilia of a HC in response to the movement of the surrounding endolymph. Cadherin-23, together with protocadherin-15, forms the tip-links and is crucial for the maintenance of the hair bundle structure[11]; mutations in *CDH23* are a common cause of pre-lingual forms of sensorineural HL such as Usher Syndrome type 1, and have more recently been found to cause forms of post-lingual HL[17].

Mutations in the *MYO7A* transcript, which codes for Myosin VIIA, also impair stereocilia function and also result in a form of Usher Syndrome type 1[5], (these forms of HL are discussed further in Syndromic hearing loss (SHL), on page 30). The physical deflection of the hair bundle, due to the displacement of endolymph fluid, occurs toward the tallest row of stereocilia and the tip links create a force causing the MET channels to open, inducing an influx of K^+ and Ca^{2+} ions into the HC body (Figure 1.5B). The exact mechanism of how this force causes the MET channels to open is still to be determined.

Upon the influx of K^+ and Ca^{2+} ions into IHC cell bodies, a temporary cell depolarization occurs which increases the cell membrane potential. This in turn induces afferent ribbon synapses at the cell base to release glutamate-filled vesicles into a synaptic cleft. The vesicles are met by glutamate receptors on afferent terminals of type I spiral ganglion neurons (SGNs) at the synaptic cleft. The main component of the ribbon synapse is the Ribeye protein, encoded by *CTBP2*. The protein otoferlin, encoded by *OTOF*, also plays a key role in the excitation process, by interacting with SNARE proteins at the pre-ribbon synapse. Deleterious mutations in *OTOF* can result in a disruption to this mechanism, causing cases of auditory neuropathy spectrum disorder and, or sensorineural HL[18,19].

Each IHC is innervated by 10-20 unmyelinated type I nerve fibres via these afferent synapses. The type 1 nerve fibres are unmyelinated until they reach the habenula perforata, where the nerve fibres align together to form the modiolus. Also at the basolateral surface of the sensory hair cells, are channels that release K+ following the

influx at the apex. Structural defects of these potassium channels can lead to impaired function; deleterious mutations in *KCNQ4* (the transcript of which codes for KCNQ4, a protein component of potassium channels in the inner ear) can result in a type of autosomal dominant non-syndromic hearing loss[11].

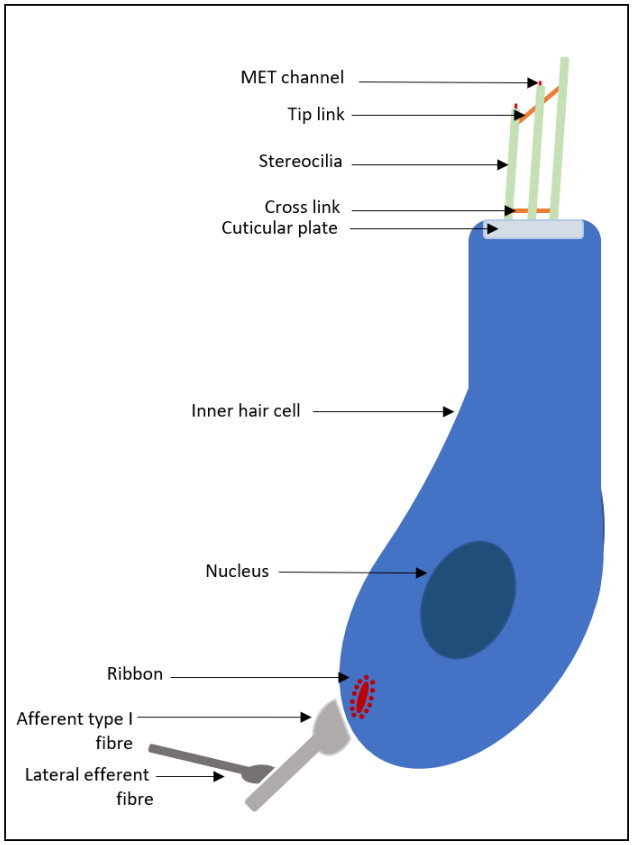

Figure 1.5. Cochlear inner hair cell anatomy.

At the apex of the IHC, lies the cuticular plate which anchors actin-rich stereocilia. The stereocilia are connected by tip links and cross links. Mechanoelectrical transduction (MET) channels are at the top of the stereocilia, and open in response to the force generated by the movement of stereocilia and connecting tip links. Ca^{2+} and K^{+} enter the cell via the MET channels, from the surrounding endolymph. Upon this influx of ions, neurotransmitter vesicles, attached to ribbons, reach the synaptic cleft. Glutamate neurotransmitters are released and met by receptors on afferent type I nerve fibres

90-95% of these auditory nerve fibres are type I[20]. The remaining population of auditory nerve fibres are type II[20], which innervate the OHCs. Similar to the IHCs, the membrane potential of OHCs increases due to the influx of K^{+} and Ca^{2+} ions. This also induces release of glutamate-filled vesicles at the synaptic cleft to meet afferent terminals of myelinated type II SGNs. This afferent signal is much reduced in comparison to the IHC afferent signal and may be involved in the medial olivocochlear reflex (an efferent feedback loop) and other functions, hypothesised elsewhere[21]. The OHCs have another, unique, physiological response to the increase in membrane potential, termed 'electromotility'

whereby the cells elongate and contract in response to the ion influx[22]. This physiological property enables the OHCs to amplify and fine-tune signals received by the cochlea, in essence extending the function of the basilar membrane.

In addition to afferent synapses, efferent synapses are present at the base of IHCs and OHCs and are components of two systems. The function the first of these systems has largely been established, it originates at the medial superior olivary complex and is thus termed the MOC system. Its role is to transmit inhibitory currents to OHCs when it is activated[23]. These currents work to reduce OHC amplifier mechanisms[7]. in response to such currents, a reduction in the motion of the basilar membrane can be observed, along with a temporary reduction in the auditory nerve sensitivity to sound[23].

The function of the second system, the lateral superior olivary (LOC) complex system, is yet to be established[24]. The system comprises of unmyelinated axons that originate at the lateral superior olivary complex and terminate as dendrites at IHC synapses[23]. Figure 1.6 is a diagram of the ascending and descending auditory pathways from the cochlea to the primary auditory cortex, including the LOC and MOC efferent systems (green arrow between the olivary nuclei and the cochlea(b)).

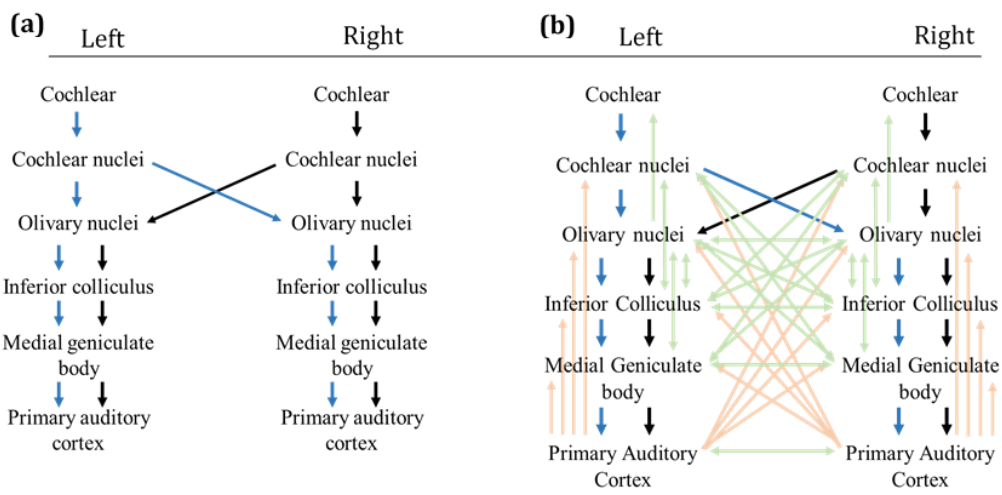

Figure 1.6. Ascending and descending auditory pathways.

(a). Ascending auditory pathway. (b). Ascending (blue/black), descending (cortical: orange; brainstem: green), and crossed (blue) auditory circuits, and crossed (green) auditory circuits. Sequential processing of acoustic stimuli takes place across the entire auditory system while parallel processing occurs mainly in the neural system. Diagram based on schematics displayed in Peterson & Renee 2019.

Returning to the afferent pathway, the auditory nerve fibres, (type I and type II) travelling from IHCs and OHCs to the modiolus, reach the central auditory processing pathway at the spiral ganglion and travel down the modiolus toward the cochlear nuclear

complex (CNC). The composition of the modiolus is also organised tonotopically. Nerve fibres that transmit signals of high frequency stimuli are located on the outside of the modiolus, while the fibres transmitting electrical stimuli from low frequency acoustic stimuli are located towards the core of the modiolus structure.

The CNC is the site where all primary auditory fibres terminate. Figure 1.7 displays the route of both the high and low frequency nerve fibres upon reaching the CNC. The low frequency fibres (originating from the apex of the cochlea) fork upon entry to the CNC. The ascending branch travels to the anterior ventral cochlear nucleus (AVCN) and the descending branch travels to the posterior ventral nucleus (PVCN) followed by the dorsal cochlear nucleus (DCN).

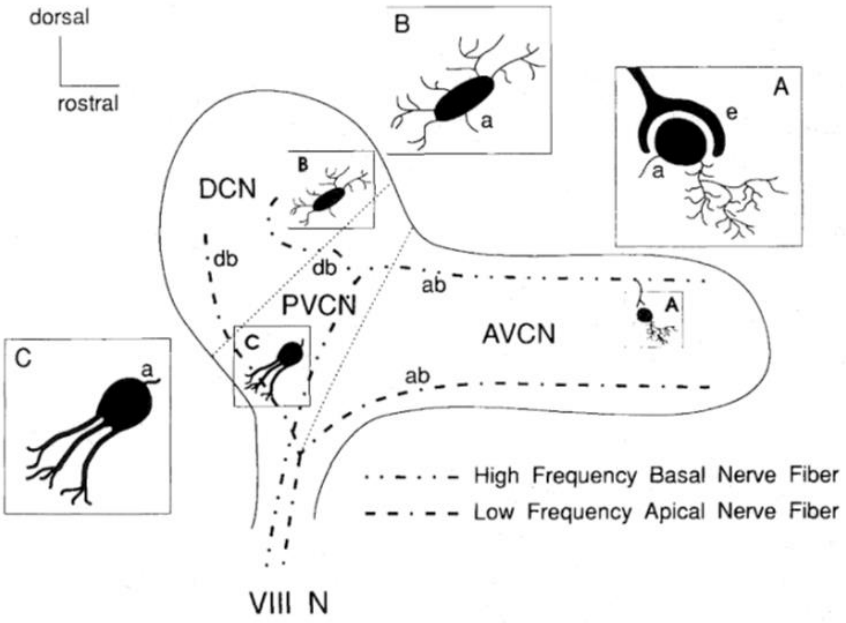

Figure 1.7. Schematic diagram of a dorsolateral view of the human cochlear nuclear complex.

Schematic diagram of a dorsolateral view of the human cochlear nuclear complex with its three divisions, anterior ventral cochlear nucleus(AVCN), posterior central cochlear nucleus (PBCN), and dorsal cochlear nucleus (DCN). Two spiral ganglion nerve fibres (VIIIN) are shown bifurcating into ascending branches (ab) and descending branches (db). Note that high-frequency (basal) fibres course in the dorsomedial portion and low-frequency (apical) fibres course in the centrolateral portion of the cochlear nuclear complex. A single spherical bushy cell of AVCN is shown in A with its axon (a) and an endbulb of Held (e) synapsing on it. A single cell of DCN is shown in B with its axon (a). A single octopus cell of PVCN is shown in C with its axon (a). Figure and legend adapted from Webster, 2005[25]

From the CNC, nerve fibres reach the superior olivary complex (SOC), primarily from the AVCN. In the SOC are three major nuclei, that are the site of the first binaural processing. The localisation and intensity differential of stimuli perceived by each ear is processed by the Medial superior olivary complex (MSO) and lateral superior olivary complex (LSO) respectively[25]. The stimuli then reach the major midbrain auditory complex; the inferior

colliculus. The cells of the inferior colliculus are organised tonotopically and have a role in the integration of multiple sound stimuli.

Sites of pathology in ARHI

Forms of hearing loss that are caused by defects or damage to structures within the inner ear or the eighth cranial nerve are termed 'sensorineural hearing loss' (SNHL). SNHL is a common condition, and the most prevalent form is age-related hearing impairment (ARHI), the focus of this thesis. The first comprehensive theory of ARHI pathology was proposed by Schuknecht in the 1960s[26]. Based on histological analysis of human temporal bone samples, he proposed that there are four subtypes of ARHI; sensory (loss/damage of hair cells in the inner ear), strial (metabolic, imbalance of the endocochlear potential), neuronal (loss of spiral ganglion cells) and mechanical (stiffening of the basal membrane in the organ of Corti). It was further theorised that each of the four subtypes would result in a distinct audiogram shape (audiogram shapes are discussed in section 1.2).

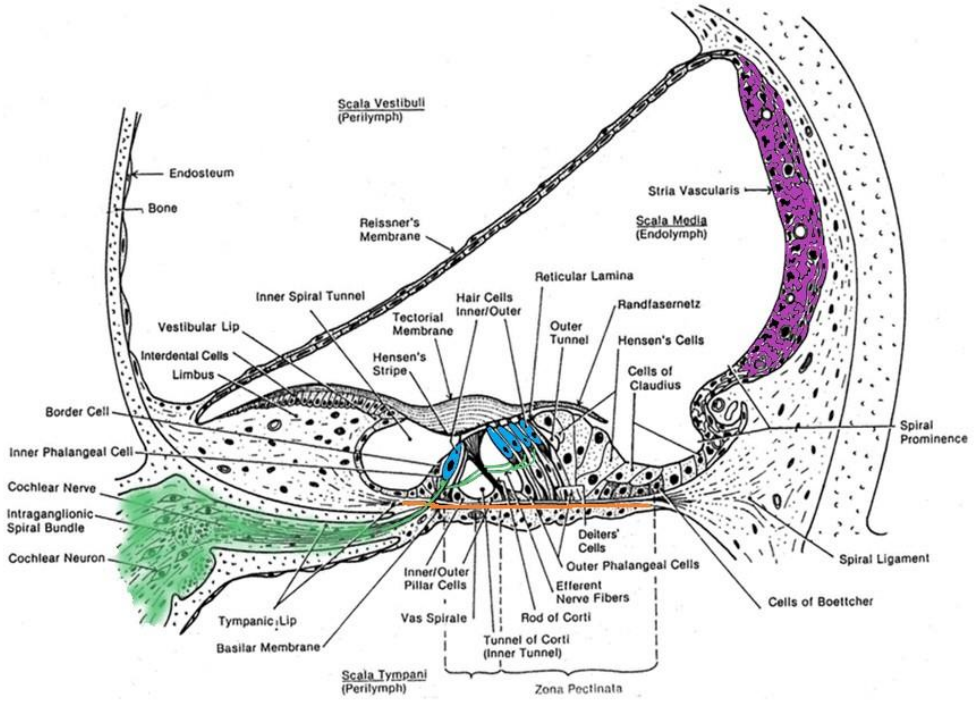

Figure 1.8. Diagram of the organ of Corti, shaded according to sites of Schuknecht's four ARHI subtypes.

Sensory, HCs shaded in blue; strial, stria vascularis shaded in pink; neuronal, neurons and spiral ganglion shaded in green; mechanical, basilar membrane shaded in orange. The original black and white diagram is from Davis and Silverman, 1970[15] and has been adapted for use here.

In 2016, the temporal bone samples were re-examined (with improved cytocochleograms) and were again compared to the audiogram measures. The work

established that the four subtypes did not significantly correlate with distinct audiogram shapes[27] and that most cases of ARHI actually present with a mixed pathology. Schuknecht's pathology subtypes described in the 1960s are therefore not commonly used in the diagnosis of ARHI, and patients' ARHI subtypes are generally not assigned based on their audiogram shapes. The subtypes do however highlight the range of distinct auditory components that contribute to the mixed pathologies of ARHI and can therefore be used as a framework for ARHI pathology.

Firstly, the sensory and neuronal subtypes do describe common physiological hallmarks of sensorineural hearing loss; the loss or damage of both sensory cells and auditory nerve fibres(ANFs),[28] Figure 1.9, and results from early research indicated that HC damage is a key contributor to sensorineural hearing loss[29,30] (as detected by a PTA). A theory was therefore proposed that HC degradation was primary to cochlear nerve degeneration[31]. According to this model of progression, the subsequent sensorineural HL is irreversible as mammalian sensory hair cells and neurons do not regenerate[32].

Subsequent studies however, have challenged this order of degeneration in the auditory tissues. A recent examination of human temporal bone specimens in a "normal-hearing" sample reported outer hair cell loss of >30% at all frequencies when comparing young (<60) and old (>60) individuals. It was also observed that in 7/11 of the subjects aged >60 years, there was a 60% loss of periphery axons compared to the 'young' group[28]. The study reports a loss of auditory nerve fibres that outweighs the observed loss of IHCs. This disputes previous findings[33,34] as it suggests that ANFs loss may be primary to or even independent of HC loss, rather than a secondary event.

Further research has supported this order of degradation, by showing that in cases of sensorineural hearing loss, damage to and the loss of synapses between cochlear neurons can occur before HC damage and long before subsequent HC degeneration[35]. This synaptic loss was not detectable on audiograms or in histological material (such as the samples used in relation to Schuknecht's work). Authors propose that this 'hidden hearing loss' may be the root cause of issues such as the inability to hear clearly in the presence of background noise; a key symptom of both ARHI and noise-induced hearing loss[35,36]. As auditory thresholds and histological studies had been the primary method of data collection in these early studies, this order would have been impossible to observe.

Repair or replacement of these neuronal synapses and a subsequent recovery of function has also recently been achieved with neurotrophin treatment in animal models[37]. This is a progression from the theory that HC degeneration was the primary pathology and that

sensorineural HL is therefore irreversible as sensory cells cannot regenerate. In addition, these recent findings demonstrate that ARHI subtypes are not as distinct as initially proposed and that the order of sensory and neuronal cell degeneration is not as straight forward as originally predicted.

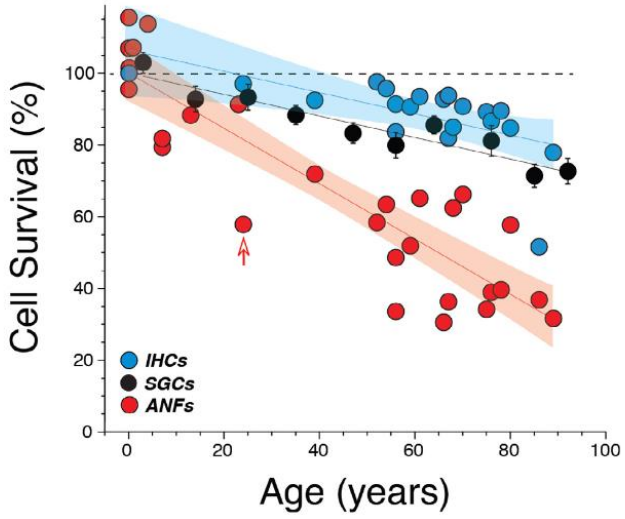

Figure 1.9. Cochlear cell survival over time.

The age-related loss of (auditory nerve fibres) ANF peripheral axons is steeper than that for SGCs or IHCs. For ANF and IHC data, each point is the cell-survival in each case averaged over all sample points within the audiometric frequencies (0.25–8.0 kHz inclusive). The SGC data are from a prior study of normal ageing individuals (n= 10 per decade of life, normalized to % survival using the y-intercept of the best-fit straight line. The ANF data are averaged over all points from 0.25 to 8.0 kHz, and also normalized using the best-fit straight line. Red arrow points to the one case exposed to a known ototoxin (cisplatin). IHC data are averaged over all points from0.25 to 8.0 kHz. IHC data were obtained as % survival and do not require normalization. The 95% confidence limits are shown for the IHC and ANF regressions. Figure and legend (legend adapted) from Wu et al., 2019.[28]

The third of Schuknecht's subtypes, 'strial', concerns dysregulation of the metabolic equilibrium of the cochlea[26]. As described, the mechanotransduction properties of the HCs and the electromotility of the OHCs is heavily dependent on precise fluctuations of the endocochlear potential. The stria vascularis is a structure that has a key role in generating and maintaining this potential via the recycling of ions in the scala media[8]. Therefore, damage to this structure, as observed in ageing mouse models of ARHI and in Schuknecht's histological samples, likely disrupts the precise transmission of the acoustic signal at the organ of Corti. As described earlier in this section, organic ions, mainly K^+ and Ca^{2+} ions, are key drivers of the endocochlear potential.

Alongside damage to the stria, metabolic function can be dependent on ion receptors and transporters elsewhere in the auditory system, such as those involved in calcium

signalling. Ca^{2+} is involved in multiple auditory mechanisms and its disruption is an established contributor to neuronal death during ageing. Null and heterozygous mice of a subunit of t-type calcium channels display a reduction in age-related cochlea dysfunction and preservation of spiral ganglion neurones. Moreover, calcium signalling could be involved in cell survival mechanisms; profoundly deaf C57 mice have been shown to exhibit an upregulation of calcium binding proteins in cochlear nucleus neurons. A putative link between ARHI and calcium signalling has also been suggested in a human population; a genome-wide association (GWAS) study highlighted a putative link (P=3.5e−7) between a genetic variant in the Collagen and Calcium Binding EGF Domains 1 gene (*CCBE1*) transcript and an ARHI phenotype[38].

There is comparatively little evidence that the fourth subtype proposed by Schuchnekt, termed 'mechanical', has a significant role in ARHI phenotypes. The subtype is based on the stiffening of the basal membrane, hindering its precise movement in response to the displacement of cochlear fluid. However, it is not unlikely that genetic variation within the population, or even individual environmental insults could result in variable function and maintenance in this structure, which could thus contribute to a sensorineural HL. A primary aim of this thesis will be to develop the understanding of these multiple pathologies which lead to sensorineural HL, specifically focusing on the aetiology of ARHI.

1.2 Hearing loss

In the population, hearing impairments and hearing loss present in many forms. In the clinic, hearing ability is generally measured by a pure tone audiometry test (PTA). PTA is considered the 'gold standard' measure for hearing assessment. A PTA provides detailed analysis of an individual's hearing thresholds independently for each ear, presented as an audiogram. The test is presented as a set of pure-tone pulses at different frequencies and intensities, the individual is asked to signal when the sound can be heard, with an overall aim to determine their ability to hear at a range of frequencies and decibels.

Diagnoses are usually given to individuals based on their PTA test results, according to thresholds determined by the World Health Organisation[39]. Normal hearing is defined as having hearing thresholds of 25dB in both ears. Those with a hearing threshold greater than 25dB in one or both ears, have a 'hearing loss' which can be classified as mild (26-40 dB), moderate (41-60 dB), severe (61-80 dB) or profound (>81 dB). A disabling HL in

adults is defined as a hearing loss greater than 40dB in the better hearing ear (https://www.who.int/pbd/deafness/hearing_impairment_grades/en/).

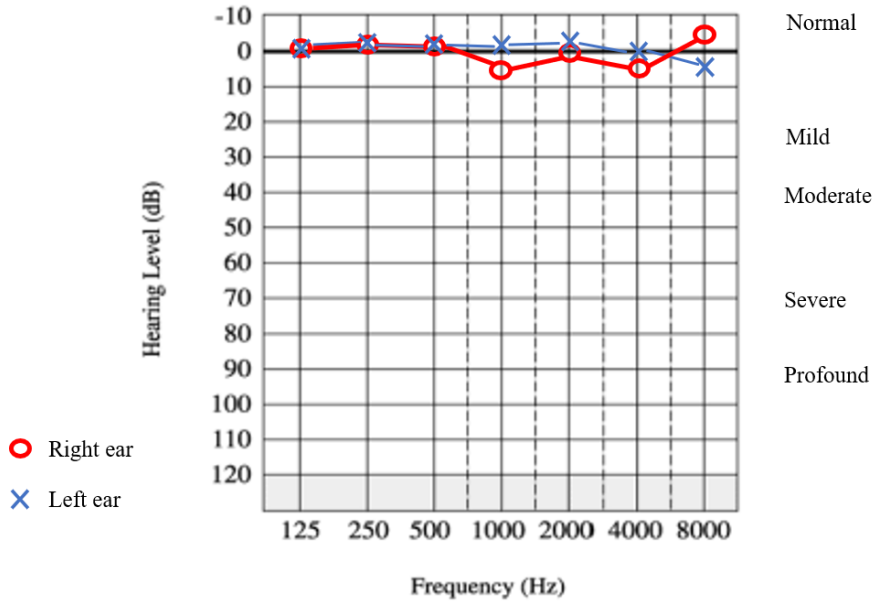

Figure 1.10. An audiogram displaying a typical pattern for someone with 'normal hearing'.

The right and left ear are tested separately at each frequency and intensity and plotted on the same audiogram. The range for 'normal' hearing is between -10 and 10 dB for both ears, at all frequencies displayed. Annotations on the right-hand side of the audiogram depict levels at which hearing abilities are described as normal, mild, moderate, severe and profound. The number of frequency thresholds that are tested can also differ between clinics.

1.2.1 Forms of hearing loss

Hearing loss can be a congenital condition (such as the pathologies caused by genes described in the previous section), a late-onset genetic condition, have an environmental cause such as acoustic trauma, or show a progressive decline in function with increasing age (such as ARHI). The various forms of HL can be characterised using different criteria as demonstrated in Figure 1.11 and Figure 1.12. Firstly, hearing loss can be divided into syndromic and non-syndromic based on how it presents clinically (Figure 1.11). Syndromic hearing loss occurs in conjunction with other symptoms representing a particular disorder. Non-syndromic hearing loss is diagnosed when hearing loss in confirmed in the absence of additional symptoms. Per 1000 newborns, 1-3 present a hearing impairment, about 20% of which are in conjunction with other symptoms (syndromic hearing impairment)[25], half of which are caused by genetic abnormalities such as those described in section 1.1.

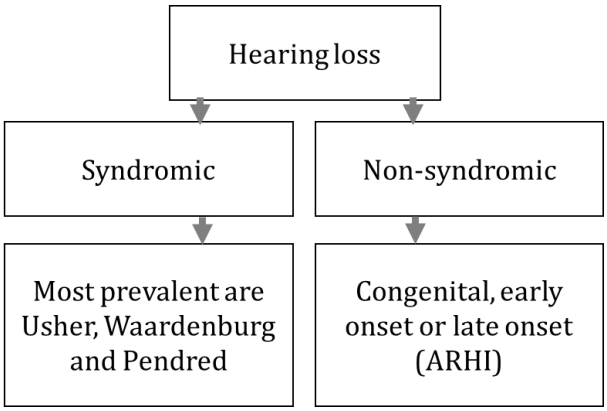

Figure 1.11. Classification of hearing loss based on clinical presentation

Alternatively, forms of hearing loss can be determined based on the location of pathology within the auditory system (Figure 1.12), also introduced in section 1.1. Conductive hearing loss, as introduced previously, describes pathologies of the outer and middle ear; the impairment affects the conduction of sound to the cochlea. These forms of hearing loss include problems with either the structure of the outer ear, the tympanic membrane or the ossicles which function to channel and amplify acoustic stimuli into the inner ear. Often the cause of these issues is either developmental (including genetic conditions that affect the formation of auditory structures) or a trauma such as a rupture of the tympanic membrane. Surgical or pharmacological are the most common types of treatment for conductive hearing loss.

Sensorineural hearing loss, also introduced earlier, consists of pathologies of the inner ear and the eighth cranial nerve. The aetiology of sensorineural hearing loss also differs between individual cases. The pathology of this form of hearing loss is due to either the development and maintenance of, or damage to and loss of the sensory cells of the cochlea, the supporting structures including the stria vascularis or the auditory nerve. It is understood that most patients present a mixed pathology of these causes such as multiple visible defects as described by Schuknecht. ARHI is generally understood to be non-syndromic, sensory neural hearing loss.

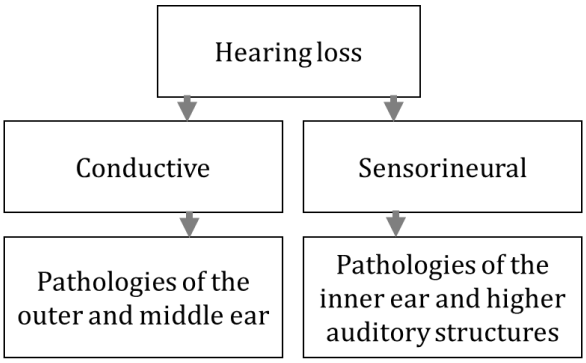

Figure 1.12. Classification of hearing loss based on the site of pathology

Non-syndromic hearing loss (NSHL)

Newborn hearing screening takes place in the majority of developed countries, which signifies the high prevalence of congenital hearing loss in the general population. Congenital forms of NSHL usually follows a pattern of Mendelian inheritance, with the majority being autosomal recessive (75-80%), presenting as a severe or profound, prelingual hearing loss. The remaining ~20% of cases are mainly autosomal dominant and resemble a post-lingual, progressive hearing loss[39]. The remaining ~5% of cases are x-linked or caused by mitochondrial mutations[39]. The most prevalent form of non-congenital NSHL is ARHI.

Forms of non-syndromic HL are highly heterogeneous (with over 100 genomic sites harbouring causal genetic variants, (https://hereditaryhearingloss.org/)). The first NSHL locus was mapped in 1988[40] via linkage analysis (Figure 1.13). Whole genome sequencing in large, clinically characterised family pedigrees has recently enabled a number of NSHL-causing genes to be identified. Genes and gene regions found to cause NSHL have a specific nomenclature. Dominant inheritance patterns are represented by DFNA*, where * denotes a HUGO gene nomenclature characterisation[41]. Similarly, recessive inheritance patterns are named DFNB*. Two identified modifier loci are named DFNM1 and DFNM2, while X-linked and Y-linked inheritance are denoted DFNX* and DFNY* respectively.

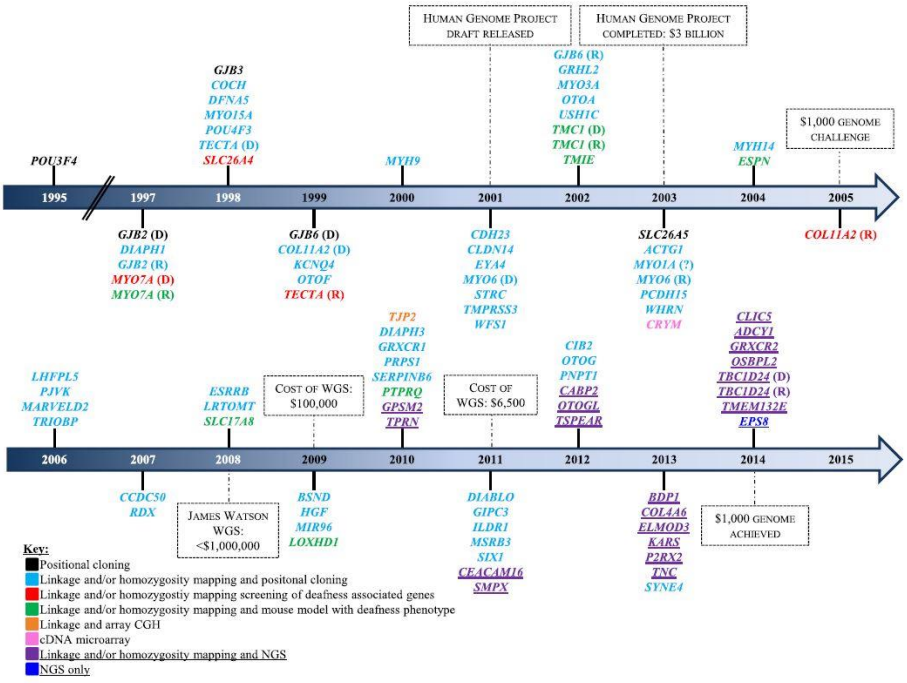

Figure 1.13. Timeline of NSHL genes identified according to method of discovery.

The timeline also detailed the major achievements in the development of NGS technologies per year[39], 1995-2015. Abbreviations: D, dominant form of HL; R, recessive form of HL; WES, whole exome sequencing; WGS, whole genome sequencing. Figure and legend adapted from Vona et al. 2015[42].

The focus of this thesis is on the genetic risk of ARHI, which is comprehensively introduced later, in section 1.3.4, yet the genetic risk loci that are involved in other forms of HL can reveal key mechanisms in hearing function, specifically those that are affected by genetic mutations. These genes can therefore be used as primary gene candidates for ARHI risk, and, or to hypothesise mechanisms involved in ARHI pathologies. Figure 1.13 demonstrates the heterogeneity of NSHL gene defects, yet a small subset of these genes cause a high proportion of known NSHL cases, specifically *GJB2* and *SLC26A4*.

Genetic defects in the *GJB2* gene (*DFNA3A*, which codes for Connexin 26)[43] have been identified to cause cases of NSHL around the world, and make up 50% of NSHL cases in populations of northern European ancestry, where different mutations and allele dosages result in a range of phenotypes[10,44,45]. Disparities between phenotypes are also observed in groups of patients that display the same *GJB2* mutations, indicative of possible dosage or epigenetic effects.

The molecular consequences of specific *GJB2* mutations have been and are currently being studied extensively. GJB2 (Cx26), in addition to Cx30, Cx29 and Cx43, are all gap junction proteins that make up the connexin channel gap junctions and, when mutated, have been found to cause various forms of HL. A hypothesised pathology of *GJB2*- related

forms of HL, is the disruption of homeostasis in the cochlea. As a component of gap junctions that line the scala media, it was theorised that GJB2 may have a crucial role in the efficient cycling of K^+ ions and Ca^{2+} signalling, and thus maintenance of the endocochlear potential and function of HC mechanoelectrical transduction. Over 100 different mutations have been identified in *GJB2* that result in a range of pathological changes. These changes predominantly affect channel formation and permeability, and are discussed elsewhere[46]. While these have been described, as of yet, no clear relationship has been observed between mutation, channel function, and hearing phenotype[46]. Of connexin proteins expressed in the inner ear, is thought that mutations in *GJB2* in particular lead to a high incidence of HL because Cx26 channels, unlike connexin channels such as Cx30, are permeable to anions and thus may have a substantial role in intracellular signalling[46,47].

Animal models have been used to further understand the pathophysiology of *GJB2* (Cx26) mutations. In one study, three strains of conditional null connexin 26 mice were generated, and all displayed NSHL[48]. Histologically, the development of structures in the cochlea of the three strains are indistinguishable from the development observed in wild type strains; HCs, supporting cells, the stria vascularis and tectorial membrane all displayed normal development. A deterioration of structures was first observed at ~P8, in the supporting cells, and at P9, the tunnel of Corti did not form in the mutant lines while full formation was observed in the wild type. By P13, wide-spread cell death had occurred at the OHCs and surrounding supporting cells and SGNs. The deterioration was first observed in the middle section of the cochlea, followed by the basal section and later the apical section[48].

A later study induced *GJB2* expression to GJB2-mutant mice via viral delivery[44]. A full recovery of gap junction function was observed in the scala media and a partial recovery of SGNs was observed. ABR thresholds of GJB2-mutant mice were not however significantly improved when compared to the wild type thresholds, possibly indicating that the viral injection occurred after the critical window for rescue[44].

Genetic alterations in *SLC26A4* are also a common cause of non-syndromic autosomal sensorineural hearing loss (DFNB4)[49]. The *SLC26A4* gene codes for pendrin, an ion transporter, and so *SLC26A4*-related hearing pathologies are therefore thought to manifest due to an ion imbalance in the inner ear, which disrupts the development and function of multiple cochlear structures[49]. While a number of causal genes have been identified in forms of NSHL (Figure 1.13), phenotypic variability is observed within family members that display NSHL. It is thought that modifier gene regions such as

DFNM1 (1q24)[50] and *DFNM2* (8p23)[50] or digenic configurations[51,52] are responsible for variation seen within some families, along with polygenic variants or genome rearrangements[42].

There is a vast range of auditory structures that are affected by the >100 genes that, when mutated, can lead to forms of HL. Examples of these were highlighted in section 1.1, Auditory function, particularly in the inner ear structures. For example, mutations in *TRIOBP* can cause a form of NSHL as *TRIOBP* has a role in anchoring of HC stereocilia rootlets, which provide the necessary rigidity for efficient mechanotransduction[16,53,54]. Mutations in *MYO7A* impair stereocilia function and so can cause HL, including forms of Usher Syndrome (see the section below)[55,56]. A further structure, the stria vascularis, can also be impaired by mutations that lead to NSHL; mutations in *CLDN14* and *MARVELD2*, which code for proteins that make up tight junctions between stria cell layers, impair the movement of ions and thus the function of the structure[11,12]. Another example, further along the auditory pathway at the pre-ribbon synapse, is *OTOF* which interacts with SNARE proteins at the synaptic cleft; mutations in *OTOF* are also known to cause NSHL[18,19].

Syndromic hearing loss (SHL)

Alternatively, hearing loss can present as one of a number of symptoms of a particular syndrome. Over 400 forms of syndromic hearing loss have been identified to date. The most common being Usher syndrome, Waardenburg syndrome and Pendred syndrome.

Usher syndrome is the most common cause of combined hearing and vision loss and is classified into three subtypes as the condition is highly heterogeneous both clinically and genetically. The three subtypes are based on the severity of symptoms and age of onset. Type I and II patients display severe to profound, and mild to severe congenital hearing loss respectively, in addition to vision impairments. The age of onset and progression of hearing loss for Type III patients is varied[49]. Nine genes have been identified as causal for USH cases. Two further genes were previously reported to cause forms of USH syndrome; *HARS*[57] and *CIB2*[58], and PDZD7 was identified as an USH modifier gene, [59] but these findings have since been questioned, see Table 1.1. The protein products of the nine genes are understood to be involved in the formation, maintenance and functioning of inner ear hair bundles[49]. USH genes listed in Table 1.1 that are expressed in HCs, are those specifically expressed in HC cytoplasm, stereocilia, tip links and cross links[60].

USH subtype	Hearing loss	Vestibular function	Retinitis pigmentosa	Identified genes
USH1	Congenital, severe to profound	Vestibular areflexia	Onset before 10 years of age	*MYO7A*[61], *USH1C*[62,63], *CDH23*[64], *PCDH15*[65], *USH1G*[66],
USH2	Congenital, moderate to severe	Normal	Onset between 10 and 20 years of age	*USH2A*[67], *GPR98*[67], *DFNB31*[67]
USH3	Progressive	Sporadic dysfunction	Variable onset	*CLRN1*[68]

Table 1.1 Summary table of the three subtypes of Usher Syndrome.

Information in the table and legend is summarised from Mathur & Yang 2014[60] and https://hereditaryhearingloss.org/usher *CIB2*[58] *and HARS*[57] have been reported to cause Usher Syndromes Type 1J and 3B respectively, but these findings have since been called in to question. It has since been shown that mutations in *CIB2* cause NSHL only[69], and the homozygous mutation identified in *HARS* is yet to be replicated and supported by functional analysis. Mutations in *PDZD7* were originally found to cause digenic USH2 with *ADGRV1*[59] but it has since been reported that mutations in *PDZD7* cause NSHL only[69].

Similarly, Waardenburg syndrome is categorised into subtypes named I-4, based on the genetic alternation that is present. Collectively, 9 genes are known currently to be involved in the development of the condition, and the majority of patients exhibit bilateral, sensorineural hearing loss[70]. The most common form, Type I, is caused by genetic alternations in the *PAX3* gene which are thought to affect the formation of a number of cochlear structures[70].

The third most prevalent form of syndromic hearing loss is Pendred syndrome. Similar to common forms of non-syndromic hearing loss, Pendred syndrome is caused by biallelic mutations in the *SLC26A4* gene[71,72]. *SLC26A4* codes for pendrin, the anion exchanger protein that works to maintain the composition and the potential of the endolymph[73]. Patients exhibit inner ear abnormalities that result in sensorineural hearing loss, and commonly an enlarged vestibular aqueduct[74]. Another gene common to both syndromic and non-syndromic forms of hearing loss is *GJB2* which, as noted, codes for Connexin 26. Mutations in this gene transcript have been linked to multiple syndromes, and, based on the nature of the mutation, can even cause multiple subtypes of the same syndrome such as seen in cases of Keratitis-Ichthyosis-Deafness syndrome[75].

While these forms of hearing loss are largely beyond the scope of this thesis, it can be hypothesised that ARHI risk loci may be common to those underlying NSHL or SHL. Therefore, genes that are known to be implicated in congenital forms of deafness are potential candidates for ARHI-risk loci and ought to be considered when evaluating

findings. Furthermore, these genetic alterations and resulting phenotypes can provide hypotheses for ARHI pathologies. While many of the genes already identified to cause NSHL and SHL cause an early-onset and severe phenotype, ARHI differs in that it is a late-onset, progressive condition. The following section discusses this form of HL in more detail and outlines the current limitations in the identification of genetic risk loci.

1.3 Age-related hearing impairment

1.3.1 Epidemiology

The focus of this thesis is on ARHI (also termed presbycusis) which, as previously introduced, is a form of sensorineural hearing impairment. It has been established that it is the most common sensory impairment in the elderly population, with the World Health Organisation (WHO) had estimated that a third of the population aged over 65 are affected by disabling hearing loss (www.who.int/pbd/deafness/estimates/en/). The true population prevalence of ARHI is challenging to determine however, because there is little standardisation across population samples, mainly with regards to which tonal frequencies are used to classify a hearing loss.

The main symptoms of ARHI are a hearing loss originating in the higher frequencies that can progress to an impairment at all frequencies[76], and impaired speech intelligibility. ARHI is clinically defined as a hearing impairment that gradually declines with increasing age and that is of a bilateral nature. As noted in section 1.2, the WHO asserted a definition of hearing loss summarised below in Table 1.2. The definition of ARHI differs between epidemiological studies, however. Studies are inconsistent in (i) the PTA thresholds tested and used to classify HL (ii) whether cases are defined by mono-or-binaural HL, and (iii) the age ranges of different study samples differ. Collectively, these factors mean that that prevalence estimates are not always comparable or even applicable to the wider population.

Pure tone threshold	Hearing ability
25dB in both ears	Normal
26-40dB in one or both ears	Mild
41-60dB in one or both ears	Moderate
61-80dB in one or both ears	Severe
>81dB in one or both ears	Profound

Table 1.2. World Health Organisation of hearing loss based on pure tone thresholds.

A disabling HL in adults is defined as a hearing loss greater than 40dB in the better hearing ear. https://www.who.int/pbd/deafness/hearing_impairment_grades/en/

In the NHANES study sample, using the WHO definition of HL, it was reported that 63% of the study sample aged >70 years had a HL[77]. In the Framingham study, the prevalence of ARHI was estimated to be 29% but was based on a hearing loss of >26dB for frequencies 0.5-2 kHz in the better ear, in subjects aged >60 years[78]. This definition further differs to that used in the Health ABC study classified a HL as 26dB, in the worse ear but which likewise assessed frequencies of 0.5-2 kHz[79]. The prevalence was estimated at 60%, but in a sample of 73-84-year-olds. Similarly, the Beaver Dam study used a sample of older age; 73% of the >70-year-olds in the sample had a hearing loss classified by a 25dB threshold loss at 0.5-2 kHz frequencies[80].

These prevalence estimates are collectively higher than in a European systematic review of multiple samples, which reported 30% of males and 20% of females to have a HL by the age of 70, and 55% of males and 45% of females to have a HL by the age of 80[81]. These slightly lower estimates may in part be due to the HL definition of a threshold increase of at least 30dB, the most conservative threshold of the studies noted. The authors of this review drew attention to the heterogeneity of the data rendering them 'unable to establish an integrative quantitative overview of prevalence rates over age and hearing loss'[81].

While there is heterogeneity in estimates based on varied pure tone thresholds, there is an additional argument for the use of alternative methods to assess the prevalence of ARHI in the population. These alternative measures may result in a more 'real life' definition of ARHI because, unlike PTA thresholds, these measures can assess an individual's speech intelligibility, (a debilitating symptom of ARHI). These alternative methods are introduced in further detail in section 1.3.4 and are comprehensively investigated in Chapter 3.

ARHI is widely understood to be a common complex trait; the condition is caused by genetic and environmental factors. These factors lead to a sensorineural HL which, as described in section 1.1, is due to either cochlear dysfunction, a type of auditory brain neuropathy, or a combination of both. The pathology is therefore highly heterogenous within the population and the condition varies between individuals with regard to age of onset, underlying pathology and rate of progression. There is currently relatively little knowledge of the specific genetic and environmental factors that impact ARHI. This limited knowledge has resulted in few treatment or prevention strategies. Despite being such a common and often debilitating condition, hearing aids are the main treatment option available, though multiple studies report less than 50% rate of uptake of hearing

aids by those with HL[82]. The following sections review the current knowledge and theories regarding ARHI development and progression.

1.3.2 ARHI risk factors

ARHI is associated with a broad range of risk factors including genetic variants, environmental exposures and health co-morbidities[83,84]. Age is the single biggest risk factor for ARHI; the risk of developing ARHI increases with increasing age. It is thought that in most cases, symptoms begin around middle age, but the age of onset varies between individuals and is not comprehensively documented at a population level. Males are also understood to have an elevated risk of developing ARHI; in a number of population samples male subjects show a higher prevalence, a faster rate of decline and an earlier age of onset than female subjects[77,85–88].

Although not all studies report these trends between male and female subjects[89], the observed differences could aid in our understanding of how the pathology manifests in different individuals[90]. One theory is that stereotypically 'male' professions have a higher risk of daily noise exposure, leading to a greater prevalence of HL and faster decline in hearing ability in male subjects. However, doubt has been cast on this theory as the observation of a higher prevalence and faster rate of decline, remains after adjusting for factors such as noise exposure[85,86,91]. Furthermore, cohorts with low or no evidence of noise exposure have reported a difference in male female prevalence[87].

An alternative theory is that sex hormone exposure may be a risk factor for ARHI. A population study found an association between hearing and a genetic variant in the oestrogen related receptor gamma gene (*ESRRG*), to be significant in females only[92]. In addition, hearing deterioration has been observed earlier in *ESRRG* KO mice compared to controls, and the deterioration was more severe in female mice[92]. Effects of progesterone have also been linked with ARHI; women taking progesterone as a component of hormone replacement therapy were observed to have poorer speech perception than controls[93]. Mice treated with different forms of HRT supported this finding; outer hair cell and overall auditory function was impaired in the progesterone + oestrogen group compared to the oestrogen only group[94]. These findings are important firstly because they indicate that hormone levels are an ARHI risk factor, and secondly that hormone therapies could potentially mediate this risk.

The risk of developing ARHI also varies between different ethnic groups. Initial observations attributed this to the environmental exposures experienced by different study samples such as industrialised vs. rural settings. More recent work however

suggests that the production of melanin may be a protective factor against ARHI[95]. Observational studies have reported that African Americans have a decreased risk of ARHI compared with White or Hispanic Americans[96]. Another study estimated that black participants had an up to 70% lower prevalence of ARHI than white participants[97,98]. A further study, in mice, indicated that melanin may be protective against SNHL following noise exposure[99]. It is important to understand why prevalence differs in different samples, as it could reveal sample-specific pathological mechanisms and direct ethnicity-specific diagnostics and treatments, promoting health equality.

Genetic risk factors – ARHI Heritability

Multiple studies have reported ARHI heritability estimates in different population samples. The estimates range between 0-79%, with the majority of estimates published prior to this work falling between 20-60%[100–105] (Table 1.3). The heritability of a trait is classically defined as the 'proportion of variation in a trait explained by inherited genetic variants.' Therefore, these estimates are evidence that genetic risk factors contribute to the variance in ARHI phenotypes observed in the population. The variation between the ARHI estimates is likely to be due to the heterogeneity between studies[38,92,106–114] (as discussed below), and indicates that a comprehensive, widely applicable estimate is yet to be determined.

There are three types of heritability estimates that capture the proportion of trait variance that is explained by inherited genetic variants. These are termed broad-sense heritability, narrow-sense heritability and SNP heritability. Broad-sense heritability (H^2) encompasses additive, dominant, epistatic and maternal/paternal effects on trait variance. Narrow-sense heritability (h^2) encompasses additive genetic effects only, and is the traditional method used for calculating heritability. For some time, twin studies were the main resource for calculating heritability. Monozygotic (MZ) and dizygotic (DZ) twins likely share the same environment, yet only monozygotic twins share the same genetic risk (excluding sporadic mutations and epigenetic modifications). Therefore, when comparing the trait prevalence between MZ and DZ twins, if the correlation is higher within MZ twins, the observed variation is more likely due to genetic factors.

Thirdly, SNP heritability (h^2_g) includes the additive effects of all SNPs specified in a model. However, the additive effects of all SNPs are not known and so two methods are generally used to calculate this term; genomic relatedness matrix restricted maximum likelihood (GREML) and linkage disequilibrium (LD) score regression. These methods that test for the proportion of variance all assume a continuous distribution for the trait

of interest and so estimates for binary traits require further adjustment to the liability scale.

The majority of ARHI heritability estimates calculated prior to this work were on twin or family samples and so most used a classical twin model to derive an estimate. However, there has been much speculation as to whether this model, which assumes equal environmental exposures, results in inflated estimates of narrow-sense heritability[115,116]. In some instances this has been disproven; a study on ECG traits compared the estimate from the classical twin model and the GREML method (which overcomes this limitation) and no inflation was observed[115]. For other traits however, such as neuroticism, a modest inflation has been detected; when similarity in environment was controlled for, a lower estimate was calculated[117]. As of yet, there are no equivalent ARHI studies that have compared the different methods, and so it is unclear whether the estimates in Table 1.3 are inflated due to this limitation.

The age range of each sample also ought to be taken into account when assessing the estimates listed in Table 1.3. Over an individual's lifetime, effects from environmental exposures accumulate. For some common complex traits this results in a decrease in the proportion of variance determined by genetic risk factors (thus a decrease in heritability with increasing age). However, this is universal for all traits, and the extent to which this applies to age-related traits such as ARHI, is not clear. To date, many ARHI study samples have been too small in size to stratify by age.

One longitudinal study however, assessed the hearing ability of participants in a Swedish male twin registry at two time points, 20 years apart, and estimated the heritability of hearing loss at each timepoint. They found no difference in the magnitude of the genetic effect between the two time points, and rather predicted that the rate of deterioration over the time period was affected by individual-environmental exposures[103]. A further study, conducted as part of the Nord-Trøndelag Health Study, also explored genetic effects on PTA thresholds, and used sibling pair correlations to estimate the heritability of hearing loss[118]. Here, the sibling correlation increased with age when calculated for sample subsets, consisting of 10-year age group samples from 20-71+ years, with the highest heritability estimate reaching 0.36[118]. Further work is required to address the reason for the wide range of heritability estimates, such as exploring the effects of age and environmental factors on the estimates.

ARHI Heritability	Study	N	Hearing measure	Age of sample (years)
0-50%[100]	4 Italian isolated populations	N = 1,682 (231, 329, 429, 693)	PTA (split low/med/high freq.)	≥40
40%[119]	Danish Twin Registry	N>2,000	Self-report	70-102
40% (unadj. for age) 25% (adj. for age)[101]	TwinsUK	N = 2,076	Speech in noise	18-87
56-70%[102]	TwinsUK	N = 1,033	PTA (PC1-PC2, PC2)	41-86
58% (aged 56-65) 47% (aged 65+)[103]	Swedish Twin Registry	N =288 N = 295	PTA (high freq)	34-99
75%[104]	Finnish Twin Study	N = 217	PTA	63-76
36-79%[105]	European population	N = 952	PTA	50-75

Table 1.3. Summary of studies that report heritability estimates for ARHI

A large number of environmental exposures are also associated with ARHI, and are discussed further in a later section entitled Environmental risk factors, p.38. These exposures can impact the heritability estimate, and possibly create heterogeneity between the studies that account for exposures and those that do not. Genes can interact with these risk factors via either increasing the risk of an exposure or influencing the response to an exposure (such as genes involved in damage repair mechanisms). The environmental exposures, and any genetic susceptibility to these exposures, can therefore influence the development and progression of ARHI.

To account for these effects, a number of the studies in Table 1.3 have adjusted for additional risk factors. For example, the study on the 4 isolated Italian populations excluded individuals aged <40 years, and, or that are diabetic or that have a history of including occupational noise exposure, ototoxic drug exposure, chromic otitis and acoustic damage. The heritability estimates were then calculated adjusting for age, sex and hypertension. Likewise, the study conducted on the European population, where heritability estimates ranged from 36-79%[105], excluded individuals that had known otologic conditions.

In addition to the adjustment or un-adjustment for environmental risk factors, and exclusion or inclusion based on medical conditions, heterogeneity between the study samples is also present in the form of the methods used to assess hearing ability. Five of

the seven studies listed in Table 1.3 used PTA thresholds to determine hearing ability. How the thresholds were averaged and thus how the hearing loss was classified varies. Firstly, three of the four Italian subpopulations studied had data available to calculate heritability estimates based on PTA thresholds, averaging 'high', 'medium' and 'low' scores. Across the three subpopulations, and across the three PTA frequencies, the (resulting 9) estimates ranged from 0-53%, yet only the 'high' PTA frequency, heritability estimate in Campara, was significant; 53% $p<0.05$[100].

Secondly, the four further studies listed in Table 1.3 that used PTA thresholds, either included only thresholds at the high frequencies[103], or derived PTA-based principle components that reflect the shape of an individual's audiogram[102]. The remaining two studies used either SNR scores from speech in noise testing, or self-report measures. Evidence of this heterogeneity being caused by different phenotyping methods, is demonstrated by the two estimates from the TwinsUK cohort. One study estimated 25-45% using SNR as the phenotype[101], and a second estimated 56-70% using a selection of PTA thresholds[102]. Although different subsets of the TwinsUK sample were used, and thus the age range of the two studies differed, subsequent work on the TwinsUK sample estimated a genetic correlation between SNR and PTA as -0.67. This is evidence that PTA thresholds and SNR scores do share genetic variance attributable to additive factors[102], yet also capture variance due to distinct genetic risk factors.

While there is variation between these estimates, and a number of studies that are limited by sample size (and thus statistical power), individually they indicate that ARHI has a heritable component, and highlight the possible effects of study-specific parameters. In order to comprehensively determine the genetic risk of ARHI on a population level, a large sample that is representative of the general population will need to be collected and an estimate calculated, possibly with adjustments for additional risk factors. These risk factors can be challenging to distinguish, especially where gene x environment interactions are present. Furthermore, the genetic risk on a population level ought not to be applied to individual patients, as the individual's risk of ARHI will be further determined by additional factors.

Environmental risk factors

Numerous environmental risk factors have been associated with ARHI, such as smoking, diet, chemical exposures, education, alcohol intake and noise exposure[120–124]. The relationships between environmental exposures and ARHI risk are summarised in Figure 1.14[120]. The effect of each exposure on an individual's risk of ARHI is poorly

understood, as are the interactions between the different exposures. The disparities between studies (displayed in Figure 1.14, column 2) regarding the effect of an exposure on ARHI risk, may be due to the heterogeneous nature of ARHI and, or the variability between the populations studied. It may also be due to the methods by which environmental factors were assessed or indeed how hearing ability itself is assessed. For example, associations with age, BMI, education and smoking have been seen to differ between men and women, at different frequencies of hearing loss[121].

Environmental factor	Comments on risk factor	References
Noise	Leisure noise causes ARHI	Clark [1991]; Lutman and Spencer [1990]
	Gunfire noise causes ARHI	
	Noise exposure increases vulnerability to ARHI	Gates et al. [2000]; Kujawa and Liberman [2006]
Chemical exposure	Toluene, trichloroethylene, styrene, xylene cause ARHI	Johnson and Nylen [1995]; Fuente and McPherson [2006]; Fuente et al. [2006]; Morata et al. [2002]
	Causes ARHI in combination with noise exposure	Rybak [1992]; Chang et al. [2006]; Fuente and McPherson [2006]; Fuente et al. [2006]; Sliwinska-Kowalska et al. [2004]
Tobacco	Tobacco use: increased risk	Mellstrom et al. [1982]; Rosenhall et al. [1993]; Helzner et al. [2005]; Cruickshanks et al. [1998a]; Itoh et al. [2001]; Nomura et al. [2005]
	Tobacco use: no effect	Brant et al. [1996]; Gates et al. [1993]; Fuortes et al. [1995]
Alcohol	Alcohol abuse: increased risk	Rosenhall et al. [1993]; Helzner et al. [2005]
	Alcohol abuse: no effect	Brant et al. [1996]; Itoh et al. [2001]
Ototoxic medication	Aminoglycosides, cisplatin, salicylate, and loop diuretics cause ARHI	Stypulkowski [1990]; Aran et al. [1992]; Boettcher et al. [1992]; Mills et al. [1999]; Chen et al. [2006]; Lee et al. [1998]; Rybak et al. [2007]; Selimoglu [2007]
Medical conditions	Renal failure	Antonelli et al. [1990]
	Diabetes	Kurien et al. [1989]; Frisina et al. [2006]
	Cardiovascular disease	Kurien et al. [1989]; Gates et al. [1993]; Brant et al. [1996]; Picciotti et al. [2004]; Torre et al. [2005]
	High bone mineral density: protective effect	Clark et al. [1995]; Helzner et al. [2005]
	Head trauma causes ARHI	Danielidis et al. [2007]; Feldmann [1987]; Fitzgerald [1996]; Rosenhall et al. [1993]
	Immune function impairment is a risk factor	Iwai et al. [2003]; Iwai et al. [2001]; Iwai et al. [1999]
Diet	Nutritional intake	Houston et al. [1999]
	Caloric restriction: protective effect	Seidman [2000]
	Caloric restriction: no effect	Willot et al. [1995]; Torre et al. [2004]
	Antioxidant intake: protective effect	Le and Keithley [2007]
Hormonal factors	Estrogen and aldosterone have a protective effect; progestin causes ARHI	Guimaraes et al. [2004, 2006]; Hultcrantz et al. [2006]; Tadros et al. [2005]
Socioeconomic status	Lower social class, no higher education is a risk factor for ARHI	Sixt and Rosenhall [1997]; Poortinga [2007]

Figure 1.14. Environmental factors associated with ARHI, from Van Eyken *et al.* 2007[120]

Three of the most well-studied environmental risk factors for ARHI are noise exposure, chemical exposure and smoking. Noise-induced hearing loss is a distinct form of hearing loss but also an established risk factor for developing ARHI. Following exposure to intense and, or prolonged sound, it is common to experience a temporary threshold shift. For repeated or extreme exposures, a permanent threshold shift can occur which signifies permanent damage to sensory cells, synapses and cochlea neurons[125].

Noise damage is detected when an elevation of auditory thresholds (a 'notch') at ~4kHz is observed (yet work by Nondhal, 2009[126] contests this classification). Both human population studies and studies with animal models have observed links between noise exposure and an elevated risk of developing ARHI[127,128], but as is evident from Figure 1.14, a number of these findings are inconsistent. One study only observed a marked increased risk of ARHI when noise exposure had occurred in conjunction with other otologic issues[128].

Furthermore, a WHO systematic review on the subject highlighted the methodological heterogeneity between studies[129]. Strict criteria meant that only five studies were included in the work, two that focused on the correlation between HL and the use of personal listening devices, two that focused on the correlation between tinnitus symptoms and use of such devises and one study that assessed both. All three studies found a positive correlation between use of personal listening devices and hearing loss, one of which displayed a positive correlation between the amount of use and HL. None of the studies displayed a positive correlation between use and tinnitus symptoms. The authors however did conclude that all evidence was 'low quality GRADE evidence', and that the data were inadequate; the combined sample was made up of n=1551 young adults and teenagers and thus difficult to measure prolonged exposure and, or persistent HL[129]. In the general population it is very challenging to quantify levels of noise exposure and the extent to which it increases ARHI risk for each individual.

Chemical exposures such as ototoxic drugs or aromatic solvents can also induce hearing loss. A common example are platinum-based drugs prescribed as chemotherapy to treat certain forms of cancer. The most frequently prescribed but also most ototoxic, is Cisplatin, which can cause sensorineural hearing loss, tinnitus and vertigo. Cisplatin forms part of a reactive aqua complex that binds DNA, inhibiting DNA replication and thus cell proliferation[130]. Cisplatin-induced damage in the cochlea, however, affects non-proliferating cells and so must induce damage via an alternative mechanism. It is thought that this mechanism is via Cisplatin binding to mtDNA which induces an increase in the production of reactive oxygen species (ROS)[76], ultimately resulting in cell apoptosis. As the cochlea is a metabolically active site, levels of mtDNA are high, and as mtDNA is histone-free, it is particularly vulnerable to Cisplatin.

Aminoglycosides are commonly prescribed antibiotics and represent another form of ototoxic drug that induces an increase in the production of ROS in the cochlea. Aminoglycosides enter the cochlea via the stria vascularis and can remain in situ long after treatment is completed[131,132]. Here aminoglycoside-ion complexes form, which lead

to the breakdown of membranes that are *in situ*. This permits the release of reactive oxygen species, to which OHCs are particularly vulnerable[133]. In the case of exposure to aromatic solvents, the molecular mechanisms that lead to a HL are less well defined, yet in population samples the risk of developing a hearing loss has been found to increase in workers that are exposed to solvents. Furthermore, this hearing loss was shown to be exacerbated when the solvent exposure is experienced in conjunction with workplace noise exposure[6]. This dual exposure was labelled as an emerging risk in the European Agency for Safety and Health at Work[134].

The link between ROS exposure and sensorineural hearing loss has been further identified in animal models and human populations. For example, uncoupling protein (UCP) genes that are expressed in the central nervous system and work as neuroprotective proteins against ROS, have been linked to ARHI. In an elderly Japanese cohort, a variant upstream of *UCP1* was associated with ARHI development[135]. Additionally, a variant in *SOD2* (that codes for a mitochondria protein that has a role in hydrogen peroxide and diatomic oxygen production) has been associated with ARHI in a British cohort[136]. These associations identified in human populations are putatively supported by in vivo studies; when subjected to caloric restriction and thus lower levels of ROS, mice displayed reduced apoptosis and decelerated cochlear degeneration[137–139].

Lastly, smoking has been an established risk factor for ARHI for decades; a study on a Beaver Dam cohort in 1998 reported that pack-years of smoking is significantly associated with hearing impairment (defined by audiometric thresholds). The study also reported that passive smokers, (defined as individuals with a household member that smokes), were more likely to have a hearing impairment than those without household members that smoke[140]. The findings have recently been replicated in the UKBB sample; Dawes *et al.* identified a dose-response effect of increased hearing impairment in both passive and active smokers[141]. Again, this is a link that has been observed but is not comprehensively understood; nicotine exposure may affect OHC function or the neuronal transmission of auditory signals[142]. Alternatively, exposure to cigarette smoke may lead to a vascular insufficiency of the cochlea[143]. Further hypothesis to explain the link is that smoking is negatively associated with a healthy lifestyle, which encompasses multiple factors that may directly impact auditory function[143].

The focus of this thesis is the genetic risk of ARHI, yet in the future it will be important to assess these in the context of all risk factors as they are not independent and the relationships between each of the factors (including genetic risk) have not been comprehensively defined. Furthermore, the work in Chapter 3 of this thesis

comprehensively assesses multiple hearing-related UKBB data fields for their suitability as surrogate measures for ARHI to use in genetic association analysis. Findings in Chapter 3 relating to the UKBB SIN test, which was used in multiple studies described in this section and the following section, challenge whether it is a reliable, stable measure and whether it really is a suitable surrogate phenotype for ARHI.

1.3.3 Associated conditions

Numerous traits have been linked to ARHI, including tinnitus, dementia, depression and medical conditions such as cardiovascular disease and diabetes[144–152]. Epidemiological studies frequently report linear relationships between the prevalence of multiple traits, which is suggestive of an association. Such associations may be due to ARHI being a risk factor for these conditions (or *vice versa*), common genetic and environmental risk factors or that they are independent traits of ageing.

Cognitive decline and dementia

Multiple studies conducted over the past 2 decades have reported an association between cognitive decline and hearing ability in old age[148–152], and in 2017, HL was defined as a modifiable risk factor for dementia in the Lancet Commission on Dementia prevention, intervention and care[153]. There are five hypotheses for the observed hearing-cognition relationship; (1) over diagnosis, where the assessment of one condition is impacted by the presence of another condition such as a cognitively-demanding hearing test; (2) widespread neural degeneration, which affects distinct areas involved in cognition and auditory processing; (3) sensory degradation and deprivation, where a hearing loss causing neuronal inactivity can result in neuronal loss and subsequent cognitive degradation; (4) cognitive resource re-allocation of neurons from cognitive functions to auditory function, resulting in a depletion in cognitive resources, and (5) social isolation/depression due to impaired hearing, which accelerates general cognitive decline[154].

The leading hypotheses is a re-allocation of cognitive load; if the input from the peripheral auditory system (outside the brainstem and auditory cortex) is degraded, more effort is required to process the signal as it reaches higher processing structures. Under this model, memory function is impaired as resources are reallocated to process the degraded auditory stimuli. The model was first proposed by Rabbitt in 1968[155] and subsequent findings have supported this hypothesis, using multiple samples under multiple methodologies[156–159].

As both hearing loss and impaired cognition are complex traits with mixed pathologies, defining phenotypes that are consistent and suitable to study in large sample sizes, is challenging. Ronnberg *et al.* studied the UKBB cohort and observed negative relationships between hearing loss and multiple visuospatial memory function tests, along with a small positive effect of hearing aid use on short-term working memory[160]. The large sample size (N=138,098), combination of hearing measures and cognitive tests performed, meant that the group could analyse subtypes of cognitive function in relation to hearing loss.

Much of this analysis supported conclusions previously published by Ronnberg in 2011, who highlighted the importance of incorporating cognitive covariates when diagnosing and treating ARHI. A study by Heinrich *et al.* in 2016 also subset phenotypes of both conditions. A number of cognitive and speech perception measures were used to identify different relationships between the specific tests [161]. Further to this, a recent systematic review on the hearing-cognition relationship concluded that hearing loss was a significant risk factor for incident dementia, but not for Alzheimers or vascular dementia[159]. Collectively, these findings imply varied relationships based on how the two conditions are measured and classified.

Multiple studies have shown hearing aid use to be positively associated with cognitive ability[162–164], indicating that hearing aid treatment may be a protective factor against cognitive decline. However, not all studies report a significant effect[164] and while the effect has been observed in studies looking at short-term use of such devices, there is currently little evidence of hearing aid use having a long-term protective effect on cognition. Furthermore, these observations may reflect the fact that more cognitively able individuals are more likely to seek and frequently use hearing aid devices. As the size of the ageing population increases, so does the burden of HL and cognitive decline. Understanding the pathology of each, and the link between the two traits, will aid in the management of both conditions.

Tinnitus

Tinnitus and hearing loss are both symptoms of a malfunctioning auditory system. A relationship between tinnitus and hearing loss has been identified[123,165], and the two traits share common risk factors including age[166,167] and depressive symptoms[168,169]. Tinnitus is the perception of sound when there is no external sound source present, and is a relatively common condition, affecting 10-15% of the population[146]. Tinnitus differs between patients by the type of sound that is perceived, the level of distress that it causes

and whether the symptoms are acute or chronic. To add to the complexity, chronic tinnitus is a symptom of a number of monogenic disorders[170].

Multiple models hypothesising how subjective tinnitus develops have been proposed. In a subset of tinnitus patients with a HL, the frequency of the sound perceived is the same frequency as their HL[171,172]. It has subsequently been hypothesised that subjective tinnitus may be a result of molecular mechanisms that increase neuron excitability to compensate for a HL[173]. Examples of these theorised models include altered lateral inhibition[173] (dysregulated inhibition of spontaneous activity at frequencies adjacent to the damaged frequency), homeostatic plasticity (an increase in neuronal gain at the damaged frequency) and stochastic resonance[174] (increased spontaneous activity via a feedback loop at the dorsal cochlear nucleus[175], to reach the neuron arousal suprathreshold).

Work to understand the relationship between tinnitus and HL is still in its infancy and is largely limited to epidemiological studies. Population studies for tinnitus, like ARHI, suffer from a key limitation; there is no objective way to measure or quantify subjective tinnitus beyond self-report symptoms and subsequent distress. Heritability estimates for tinnitus vary from 11%[176] in the first population study, to 68% in the male subset of a Swedish twin cohort[177], to the most recent estimate of 40%[178]. Identifying common and distinct risk factors such as genomic risk loci for both traits would progress the current knowledge of how the traits develop individually and in conjunction with one another.

Depression

A recent, systematic review and meta-analysis using the literature on the relationship between hearing loss and depression, concluded that there was a small but significant association between hearing loss and greater odds of depression[140]. Of the 44 studies included in the meta-analysis, 40 observed a positive association between hearing and depressive symptoms. Such studies have reported associations between depression and anxiety with hearing loss at high frequencies[179], a higher incidence of depression among hearing loss sufferers, and also a lower incidence of depressive symptoms in hearing aid users[180,181].

A lower incidence of one condition, depressive symptoms, following the treatment of another condition, ARHI, with hearing aids, is suggestive of the latter being a risk factor for developing the former. If this hypothesis could be proven by identifying an underlying biological mechanism, improving treatments and prevention strategies for

ARHI could reduce the burden of depression on public health and subsequent economic resources.

Conversely, another theory to explain the relationship between the two traits, is that HL hinders social interactions, and that this could lead to an onset of depressive symptoms. A recent study however has shown that psychosocial factors, such as participation in, and access to, social activities may not be responsible for the association, as adjustment for these factors did not attenuate the association[182]. Consistent with this finding, a second study in 2018 reported a possible common neural degradation that is in-part causal to both hearing loss and depression, based on neuroimaging of the limbic system, frontal cortex and auditory cortext[182].

Approaches to disentangle relationships between ARHI and associated traits

The relationship between any of these traits and ARHI can be disentangled by a number of methods. Longitudinal analysis can reveal epidemiological trends of disease onset, progression and association between traits, though longitudinal studies require decades of foresight and rely on participant retention. Studies to identify common biological mechanisms can be more informative and efficient to conduct. For example, common biological mechanisms can be discovered by establishing whether two traits share common genetic risk factors. Techniques such as genetic correlation analysis are therefore a more practical approach to disentangling these relationships.

Genetic correlation analysis between traits can assess the level of common genetic variance and can be performed with distinct trait-specific cohorts, thereby omitting the need for novel large-scale longitudinal data collection. Mendelian randomisation can be used to explore whether one trait is likely to be a modifiable risk factor for another, based on the nature of the shared genetic variance[183]. These techniques are especially valuable as they reduce the effects of confounding factors, which are difficult to omit in observational longitudinal studies[184]. Prior to this work, these techniques have not been used to disentangle the relationship between hearing and cognitive decline, tinnitus or depression.

1.3.4 Strategies for defining ARHI in human populations

Valid epidemiological and genetic analysis of complex traits relies on defining and ascertaining a robust phenotype. In quantitative traits this often consists of a measure by which an individual is classified by trait severity. Qualitative traits are assessed with a case control design with respect to the trait of interest. Pure tone audiometry (PTA),

signal to noise ratio (SNR) testing and self-report measures have all been used to define population prevalence and severity of ARHI.

Assessing hearing ability: Pure Tone Audiometry

PTA tests, as discussed in Section 1.2, are the gold standard measure of hearing impairment; an individual's ability to hear at a range of frequencies and pitches is assessed and presented in the form of an audiogram. ARHI generally presents as a bilateral deterioration of hearing function that originates in the higher frequencies, as depicted by the downward sloping audiogram in Figure 1.15. To date, PTA testing is the most common ARHI phenotype used in most epidemiological and genetic studies of ARHI. However, PTA tests are not optimal or indeed practical for large-scale population studies as they require expensive, specialist equipment and trained staff. In addition, the test does not encompass a measure for distinguishing and interpreting sounds which are presented alongside background noise, sometimes termed 'hidden hearing loss', a common symptom of ARHI[28,185].

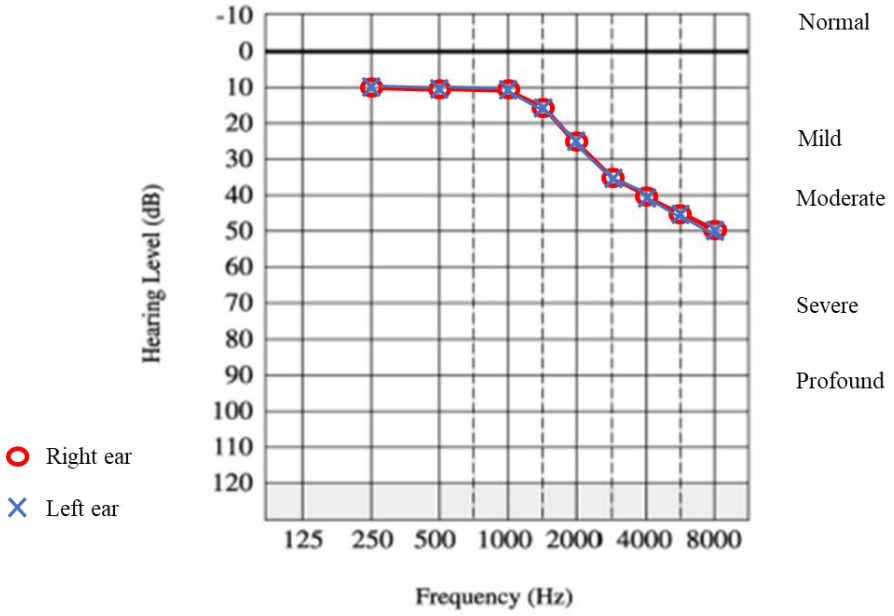

Figure 1.15. Example audiogram for a patient with age-related hearing loss.

A hallmark of ARHI is a downward sloping audiogram for both the left right ears as it signifies a deterioration in hearing in the high frequencies and that the deterioration is bilateral. Hearing loss is classed as asymmetrical where there is greater than a 10dB difference between ears at the same frequency. Annotations on the right-hand side of the audiogram depict levels at which hearing abilities are described as normal, mild, moderate, severe and profound. Not all cases of ARHI will exhibit this downward sloping audiogram and there is no consensus to define ARHI by individual or grouped thresholds. The number of frequency thresholds that are tested can also differ between clinics.

Assessing hearing ability: Speech in noise testing

Speech in noise (SIN) testing determines an individual's ability to detect and process sounds at different frequencies with varying levels of background noise. The test produces a speech reception threshold (SRT) for each individual, indicating the level at which they can correctly interpret 50% of the speech presented to them. The SIN test has been proven as a useful surrogate for a PTA measure in the study of ARHI[102] and is likely to be a more realistic measure of how ARHI impacts an individual's ability to hear in daily life due to the incorporation of background noise. It is also amenable to genetic studies with large cohorts as it can be completed online or via telephone without additional expensive equipment or trained staff present. Different tests use either sentences, words or numbers as speech material. Currently, however, there is a lack of large cohorts with both SIN and genetic data available and testing protocols vary between studies[102,113,186]. The UKBB however, has recently released SIN and genetic data on a sample of >160,000 individuals; magnitudes larger than other samples with SIN and genetic data.

SIN tests were initially developed to provide an accessible, objective measure of hearing impairment, with the aim to increasing the diagnosis of hearing impairment and the subsequent uptake of amplification devices. SIN tests are a more accurate measure of 'real world hearing impairment' as they mimic real-life situations more accurately than PTA. The standard speech in noise test was devised by Plomp, 1976[187] and consists of a series of sentences played against a background noise. Participants then relay the sentence to be best of their comprehension. The test has an adaptive procedure; if the participant relays the sentence correctly, the dB ratio of the speech to background noise is altered such that the background noise level is increased. This continues until the intelligibility of the sentence reaches 50%. The ratio of speech to background noise at this 50% threshold level is the participant's 'speech reception threshold' (SRT).

The test has since been adapted for use via different platforms in multiple countries. In 2004, the first digit-triplet test (DTT) for use by telephone was devised by Smits[188]. This differs to the SIN test devised by Plomp, as numbers (digits) are used in place of sentences or words. While the DTT has its advantages, it is less sensitive to cognitive function than alternative forms of SIN testing[122,189,190]. Triple digit SIN tests vary in the numbers selected to use, speaker that is employed, concatenation method, noise employed, scoring method and the tracking rule[191–194]. The UKBB SIN was based on the original Dutch online SIN test, first adapted for use in the UK by Action on Hearing Loss and further adapted for use in UKBB. The three key differences between these three tests (except the language used Dutch/English) are: (1) the tracking rule, (2) the number of

digit-triplets presented during the test and (3) the number of SNR measures used for the final SRT calculation (Table 1.4).

Speech in noise (SIN) test	Triplets played	No. of triplets for SRT	Tracking rule
Dutch online Speech-in-noise	23	23	1-up 1-down (2dB)
Action on hearing loss (used by TwinsUK)	26	26	1-up 1-down (2dB)
UKBB Speech in noise test	15	8	1 up 1 down (4dB) for triplets 1-7, 1 up 1 down (2dB) for triplets 8-15

Table 1.4 Summary of the adaptations to the original Dutch SIN test firstly for use by AoHL and secondly for UKBB.

No. of triplets for SRT, the number of SNR ratios used to calculate the SRT value. The SRT is the mean SNR of these triplets.

The SIN-PTA relationship

Cohorts with both SIN and PTA data have been used to test whether the SIN test is an adequate surrogate for PTA as a measure of hearing ability. In the first study N=39 (78 ears), correlations were calculated between SRTn and PTA thresholds 0.5,1,2 ($r^2 = 0.732$) and SRTn and PTA thresholds 0.5,1,2,4 ($r^2 = 0.770$). The sensitivity and specificity of the SRTn test to PTA measures were also calculated as 0.75 and 0.91 respectively for PTA thresholds 0.5,1,2 and 0.79 and 1 for PTA thresholds 0.5,1,2,4[188]. The validation of the SIN test has also been tested in the TwinsUK cohort; N=351 female northern-European participants have completed both SIN and PTA tests. The correlation between SIN scores and PTA was r = 0.62, while the sensitivity and specificity calculated to be 89% and 80% respectively for a threshold of -9.25dB SNR when compared to PTA[102]. Furthermore, the genetic correlation (representing shared genetic variance attributable to additive factors) between the two traits in the TwinsUK cohort was calculated as -0.67[102].

The two studies used slightly different SIN tests (Dutch telephone test and Action on Hearing Loss web-based test), yet both demonstrate high correlations, sensitivity and specificity to PTA measures and thus support the use of SIN as a surrogate measure. In addition, the difference seen between the two tests (such that the sensitivity and specificity is not 100%) is not necessarily a limitation. Unlike the PTA test, the SIN test assesses a participant's ability to process speech; patients that have difficulty deciphering speech can present with a normal audiogram. So, while validating the SIN

test against the PTA test is important, the differences between the two tests may have equal or even more importance when assessing ARHI.

Self-report hearing measures

Self-report measures are an increasingly common and practical method for the collection of data for a number of traits, on a large scale. In the analysis of complex traits, a larger sample size generally increases the statistical power. Although self-report measures can be subjective, genetic associations with ARHI have been found using this approach[113]. Self-report hearing measures may be simple questions regarding an individual's hearing ability under different conditions, use of hearing amplification devices or any previous diagnoses of hearing-related problems and related symptoms.

Several studies have reported the prevalence of hearing loss as measured by self-report status. These are summarised in a systematic review which was conducted with the aim of estimating the prevalence of presbycusis in the European population aged >60 years. The review reported on seven studies with self-report data (in addition to studies with PTA data)[195]. Prevalence estimates in the seven studies ranged from 8.1% (>60 years, Italian population)[196] to 40.3% (>= 75 years, UK population)[197]. The variation in these estimates may be due to the different questionnaire measures used and may also reflect the varied demographics of the population samples.

Two publications have reported on self-reported hearing aid use. The first summarises the correlation between frequent hearing aid use and SRT score[123] and the second explores whether the use of hearing aids is associated with better cognition[162]. The first study reported that only 21.5% of adults that had 'poor' hearing as defined by the UK Biobank (UKBB, introduced in section 1.4 and used extensively in this thesis) SIN test wore a hearing aid 'most of the time'. The emphasis of the study was the relatively low uptake of hearing aids by those with a hearing impairment[123]. The second study reported a positive association between hearing aid use and cognition, independent of social isolation and depression[162].

The relationship between self-reported hearing ability and PTA and SIN measured hearing ability

Four studies have collected both self-report and PTA data, and the prevalence estimates within studies differed by 7-16%[198–201]. A further study in 2016 reported similar findings; a cohort of 1669 participants aged >70, self-report measures and PTA scores were assessed. The self-report was not found to be truly representative of hearing

impairment as measured by PTA, indicating that the two measures may represent different forms of hearing impairment in this cohort, and possibly in cohorts that have been studied previously[202].

A study in 2015 describing the relationship of functional hearing and depression in the UKBB cohort reported a small (0.65dB) but significant difference between SRT_B scores of those that reported self-reported hearing difficulty and those that did not. The study also reported a small but significant difference in SRT_B score (0.51dB) for those reporting difficulty and no difficulty hearing in background noise[195]. However, a further study reported inconsistencies between participant perception of hearing ability and SIN measured hearing ability; 25-30% of participants that measured poor hearing (based on their UKBB SIN score) reported no difficulty with their hearing, and 25-30% that measured normal hearing reported hearing difficulties[195].

Based on the studies discussed in this section 1.3.4, it is clear that there are different ways to measure hearing impairment in population samples, each with individual merits and limitations. It would be informative to assess the prevalence of hearing impairment as measured by questionnaires, PTA thresholds and by SIN scores to determine which phenotype is most representative of ARHI, and to better interpret previous studies that have used each method of assessment.

1.4 ARHI genetic risk

ARHI is predicted to be genetically heterogeneous, indicated both by ARHI genetic studies to date[38] and the fact that 150 genomic loci are associated with non-syndromic hereditary hearing loss alone (https://hereditaryhearingloss.org/). Identifying genetic risk factors for this trait would aid the understanding of its pathology and ultimately enable targeted treatments and prevention strategies. ARHI heritability estimates range between 30-70%[100,101,103–105,119] (as discussed previously in Genetic risk factors – ARHI Heritability, page 35) but to date there is little understanding of which genetic loci have a role in pathogenesis, and what their roles are. ARHI has been studied in human population studies, primarily with GWAS and candidate gene approaches, yet only five genomic loci have been significantly associated with ARHI, of which two have been replicated in independent samples[109,111–113].

1.4.1 Identifying genetic risk variants

Several methods are commonly used to detect trait-specific genetic risk variants. In terms of ARHI, a number of these methods have been used to estimate genetic risk factors and variant associations with the phenotypes discussed previously. To determine

the variance in phenotypes that is explained based on genetic factors, a measure of heritability can be calculated (section 1.3.2). To identify individual genetic regions that likely cause this variance, techniques such as linkage analysis, candidate gene studies, genome-wide association studies and genome sequencing are applied, as discussed below.

Linkage analysis

Linkage analysis is a traditional approach used to identify chromosomal regions that harbor genetic disease markers. The technique makes use of the laws of genetic recombination; markers in close proximity on chromosomal regions are more likely to be inherited together. Family pedigrees are therefore used to identify chromosomal regions that cosegregate with a disease or trait of interest. Logarithm of odds scores are calculated to assess the probability that the observation is due to linkage rather than chance. Prior to the development of methods such as GWAS, linkage analysis was the optimal method to discover regions of the genome that are implicated in disease.

A linkage analysis study was conducted on a group of ARHI cases presenting with high frequency hearing loss. The *DFNA5* gene had been identified in non-syndromic sensorineural, progressive hearing loss in multiple families of different genetic backgrounds[203–205] and was therefore a strong candidate gene for ARHI. Analysis was conducted using a number of microsatellite regions on *DFNA5* but no evidence of linkage was observed[93]. Following this, a genome-wide linkage study for ARHI was performed with a sample of 1789 individuals from the Framingham study, derived from 328 extended pedigrees[206]. At the time, heritability of ARHI had been estimated as 40% in a classical twin study from the Danish twin register [119], while the linkage study provided broadly similar results with heritability estimates of 31% and 38% for age-adjusted low- and medium pure tone average frequencies respectively. Six chromosomal regions were identified as having suggestive evidence of linkage, four of which contained genes known to be involved in forms of deafness, three of which being implicated in Usher syndrome. However, while this study may indicate the role of congenital deafness genes in adult progressive hearing loss, no truly statistically significant findings were reported.

Linkage analysis has also been performed with the self-report hearing measures. One study used a cohort of elderly male twin US military veterans. Four hundred markers were studied, and a region of suggestive linkage was identified on chromosome 3, in the same region that *DFNA18* is located. This provides the first evidence to suggest that *DFNA18* has a role in general hearing impairment in the ageing population[207]. A further

linkage study was performed, and although no significant associations were identified, variation in a region on chromosome 8 was correlated with a concave audiogram shape[208].

In keeping with many other common complex traits, there has been limited success in identifying ARHI susceptibility loci via linkage analysis, due to its' inherent limitations. The power to detect linkage via pedigree analysis relies on the presence of highly penetrant, rare variants such as those in Mendelian or oligogenic disorders[209] yet common, polygenic traits such as ARHI are caused by multiple, lower penetrance variants; the 'common trait common variant hypothesis' (Figure 1.16). A second limitation of linkage analysis is its limited power of resolution; the regions denoted by the widely spaced markers - usually microsatellites - are large and may contain hundreds of possible candidate genes. Where a region is significantly associated with the trait of interest, further analysis is then necessary to narrow down the genomic region of interest and identify the functional genetic variant or alteration.

Candidate gene analysis

A second method, candidate gene analysis, tests for association between a trait and a gene of interest. This approach may be used to identify a causal SNP or genome variant when there is prior knowledge that a gene, biological mechanism or pathway is associated with the trait of interest[210]. When genotyping costs were relatively high, in the early 2000s, this was the most cost-effective approach to use, as it only necessary to genotype the genome region of interest.

As noted earlier, *DFNA5* was considered a strong candidate gene for ARHI as it causes some forms of familial deafness. With the same data as used in the linkage analysis study, a case-control association study was conducted with satellite markers from the *DFNA5* candidate gene region, yet no evidence of association was found[93]. Another candidate gene study used samples from 7 different populations and tested for association between ARHI and *GSTT1, GSTM1* and *NAT2.* These genes were selected on the basis that they have a role in defence against reactive oxygen species, a putative pathogenic mechanism in ARHI. In the combined European samples a significant association was observed with *NAT2*6A*, while associations with *GSTT1* and *GSTM1* were observed in the Finnish subsample[211]. In 2008 a more comprehensive candidate gene study of ARHI was performed using 768 SNPs across 70 genes. The 70 genes were selected as they are genes known to cause monogenic forms of hearing loss genes in both mice and men. Here, *GRHL2* was found to be associated with ARHI; one SNP reached the multiple-testing p

value threshold of association and surrounding SNPs showed suggested association, with a concordant direction of effect in all nine contributing cohorts[212].

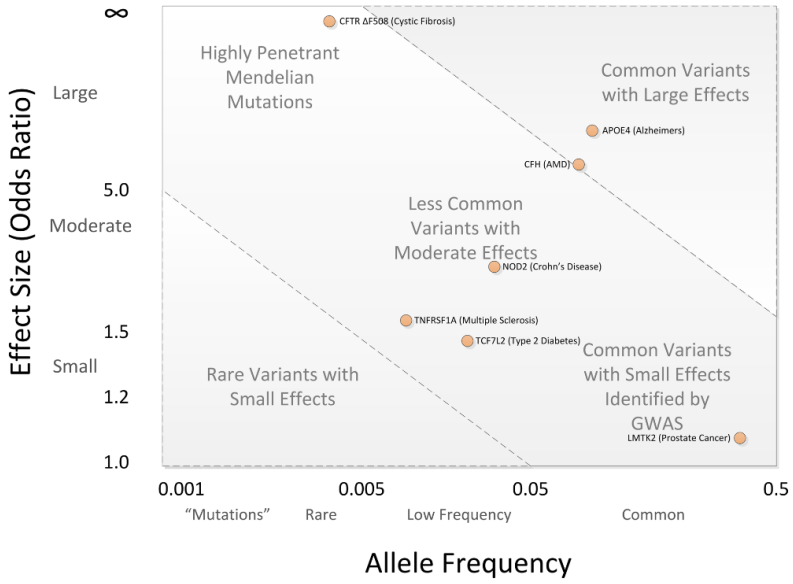

Figure 1.16. Diagram to demonstrate the common variant common trait hypothesis.

Spectrum of Disease Allele Effects. Disease associations are often conceptualized in two dimensions: allele frequency and effect size. Highly penetrant alleles for Mendelian disorders are extremely rare with large effect sizes (upper left), while most GWAS findings are associations of common SNPs with small effect sizes (lower right). The bulk of discovered genetic associations lie on the diagonal denoted by the dashed lines. Figure and Legend from Bush & Moore 2012[213].

The most comprehensive, ARHI-specific, large-scale candidate gene study to date on normal hearing was conducted using a multistep approach[203]. Nineteen candidate genes that had previously been reported in ARHI association analyses (yet were not genome-wide significant), were examined for gene expression in mouse cochlear tissue. Gene expression in the cochlea was observed for 12 of the 19 candidate genes. Of the 12, 9 were then replicated in independent samples. A relationship was also identified between specific audiometric patterns and variants in 7 of the replicated genes; *CDH13, GRM8, ANK2, SLC16A6, ARSG, RIMBP2* and *DCLK1*.

The gene *GRM7* contained the first variant to be identified by association analysis, in 2009, in a population sample comprising cases and controls collected in 8 centres from 6 European countries (n=1700 total). Replication was attempted for the most highly associated variants, in a sample n=138, and the variant in *GRM7* variant was replicated[109]. *GRM7* has been the focus of several subsequent candidate gene studies. The first study which aimed to replicate this finding was performed on a European-American population (n=687), and used a range of hearing measures in an attempt to

better define different pathologies underlying a ARHI phenotype[106]. Six phenotypes were defined and tested for association with three variants of interest from *GRM7*; rs11928865, Haplotype 6 and Haplotype 7. The study confirmed an association between variation in *GRM7*, PTA frequencies and SRT scores.

A later study assessed whether *GRM7* variants were associated with ARHI risk in a Han Chinese population sample[108]. Individuals were assigned to phenotype groups based on the shape of their audiograms, using K-cluster analysis, following Schuknecht's model of the four audiogram patterns of ARHI[26,214] subtypes. Significant associations were identified for rs11928865, a SNP in intron 1 of *GRM7* for the audiogram phenotypes 'abrupt loss' and 'sloping shape', but not for 'flat shape' and '8kHz dip'. This work was another indication that different risk genetic factors may associate with distinct audiogram shapes. The latest study to assess *GRM7* risk variants in an ARHI population sample tested 4 SNPs for association, and found one to associate in a dominant pattern[215]. Two studies have analysed *GRM7* risk variants in relation to traits linked to ARHI; noise-induced hearing loss in a Han Chinese sample[216] and tinnitus and ARHI in a Portuguese sample[217]. Furthermore, a number of subsequent GWAS reported summary statistics for *GRM7* SNPs included in their analysis[38,107].

The multiple findings of an ARHI-*GRM7* association are promising, however further validation is required for multiple reasons. Firstly, associated SNPs are located in different regions of the *GRM7* gene and were identified using differing phenotype measures in diverse ethnic samples. Secondly, the initial GWAS study did not report a significant association and the latest association studies, with greater statistical power, have failed to support an association with *GRM7*[113].

Interpreting these findings is complex, and a pathogenic mechanism of *GRM7* in ARHI is yet to be established. Furthermore, as the candidate-gene approach is hypothesis-driven, it is susceptible to false positive results. Evidence based on work in other traits indicates that candidate gene association studies have provided many false positive results, demonstrated by more recent, highly powered GWAS[218]. The recent development of very large population samples could lead to a rapid expansion in studies of ARHI. Due to the falling costs of genotyping and the advent of the GWAS-era and thus hypothesis-free methods, candidate gene analysis is now a less favourable method to use for discovering novel genetic associations with traits of interest.

Genome-wide association studies

The last decade has seen a rapid increase in the number and complexity of methods to uncover genetic risk variants associated with common complex traits. The falling cost of genotyping, the increase in large-cohort data collection and the development of high-powered analysis platforms has led to wide-scale usage of GWAS to uncover genetic determinants of polygenic traits. The hypothesis-free approach tests for association between a trait of interest and up to billions of common genetic polymorphisms. Over the past decade, GWAS has enabled the discovery of genetic risk variants for a whole host of polygenic traits. For ARHI however, there are still very few genes that are known to be involved in pathogenesis, and heritability estimates, as noted, vary greatly depending on the characterisation of the condition, cohort demographics and analysis method.

For an individual trait, the success of GWAS is determined by four main factors; the number of trait-associated loci that segregate in the population, the genetic architecture of these loci (including effect size and frequency), the size of the study sample and the heterogeneity of the trait[219]. For ARHI, prior to the work presented in this thesis, the number of trait-associated loci segregating in the population, and the architecture of these loci was largely unknown.

The initial GWAS studies on ARHI used cohorts in the hundreds of individuals while studies in the past ten years have featured samples in the thousands and have increased genome array densities and utilised vast imputation panels. Although this has improved the power to identify associations, there is still a distinct lack of replicated associations from ARHI GWAS. This is likely due to the polygenic and multi-factorial nature of the trait, and the differing cohorts and study designs; previous ARHI GWAS differ in sample size, study population, SNP coverage and phenotype definition (Table 1.5).

The first GWAS on ARHI was published in 2009. The sample included 1,692 participants, classified either case-control based on Z-scores derived from PTA frequency thresholds[109]. The study did not report any genome-wide significant associations, yet the lead suggestive association, as noted in the section above, was a SNP in the *GRM7* gene transcript. The association replicated in an independent sample and subsequent gene expression analysis and immunohistochemistry of its protein product, metabotropic glutamate receptor 7, was observed in mice cochlea and in a single human cochlear sample.

The following year, a second GWAS was published, and defined ARHI , using principal components (PCs) derived from PTA thresholds[107], resulting in a quantitative phenotype.

Again, no significant associations were identified in the sample (n=352), but the most highly associated SNP was reported in the gene *IQGAP2*. A further GWAS also used quantitative phenotype, based on PCs derived from PTA thresholds. This study meta-analysis of samples from across Europe and so had larger sample size (n=3,417). No significant associations were identified, but authors highlighted four candidate genes; *DCLK1, PTPRD, CIMP* and *GRM8*[114]. Replication was not conducted in the original study but the association with *GRM8* was later supported in a subsequent GWAS[38].

In 2013, Nolan *et al.* studied three cohorts (n=3,900) using a quantitative phenotype derived from logistic regression analysis of PTA thresholds. The focus of the study was *ESRRG* and identified an association in females but not in males, in two of the three study cohorts[92]. The link was supported by functional work that was conducted in *ESRRG* knock out mice. A year later Fransen *et al.* performed a GWAS and attempted replication of all previous associations, with a phenotype derived from PCs of PTA thresholds (N = 2,161)[38]. None of the previous findings were replicated and no new significant associations were found. At this stage, none of the five ARHI association studies had revealed any significant associations with risk variants, and there was little replication of suggestive findings between studies. The authors highlighted suggestive associations with *ACVR1B* and *CCBE1* and concluded that ARHI is likely to be 'highly polygenic in nature' and that the trait is caused by multiple variants of small effect sizes, which are 'undetectable in a modestly powered GWAS'[38].

Later in 2014 two studies published the first genome-wide significant associations between genetic risk variants and ARHI. Firstly, a SNP in the salt-inducible kinase 3 *(SIK3)* gene transcript was identified in a meta-analysis of samples (overall n=4,939), where the phenotypes were derived PC analysis on PTA thresholds. This finding is yet to be replicated in an independent sample, yet Sik3 expression was observed in the stria, hair cells and spiral ganglion neurons in P0-P5 mice[111]. The second study identified significant associations near to the *SLC28A3* and *PCDH20* gene transcripts, and reported replications at nominal levels of significance[112]. In 2015, Fransen *et al.* published a second GWAS with a cohort that had been used previously in a candidate gene study. No significant novel associations were identified, and replication was not achieved for any previous findings of ARHI-associated genes[38].

Study	Phenotype	Genotyping array / no. of SNPs	Sample size (analytical method)	Ancestry of samples	Genes identified P<5E-08
Friedman *et al.*, 2009[109]	Z-scores derived from PTA thresholds 2,4,8 kHz (high frequency, most affected in the elderly). Use only the 'best ear.	Affymetrix 500K GeneChip® / 506,627	N = 846 cases and N = 846 controls	European	-
Van Laer *et al.*, 2010[107]	PC analysis of PTA thresholds. PCs 103 captured 80% of the variation. PC1 correlated with all frequencies. PC2-3 are 'shape' variables. PC2 is a measure of the slope of the audiogram, PC3 measures the concavity of the audiogram.	Affymetrix GeneChip Human Mapping 100 K array pair (Santa Clara, CA, USA) / 83,381	N=347 (quantitative)	Finnish Saami	-
Girotto *et al.*, 2011[114]	(1) PTA low, medium and high (2) 7 different thresholds 250-8k Hz (3) PTA threshold PCs 1-3 (shapes as in Van Laer *et al.*)	Illumina 370 k platform or Affymetrix 500K. Imputed to the 2.5M HapMap CEU SNP set v22 / ~2,200,000*	N = 3417 (meta-analysis)	European	-
Nolan *et al.*, 2013[92]	PTA at 1kHz and 4 kHz 0.25-4 kHz	ABI TaqMan SNP genotyping assay, Illumina Infinium Human Hap550 array / 555,164[a]	N = 3900 (quantitative)		-
Wolber *et al.*, 2014[111]	PC analysis of PTA thresholds (shapes as in Van Laer *et al.*)	Illumina 370k chip, Affymetrix 500k array, Illumina HumanHap300 Bead Chip, Illumina HumanHap610 Quad Chip / >2,300,000*	N = 4939 (meta-analysis)	European, Silk Road	*SIK3*

Vuckovic *et al.*, 2015[112]	PTA 0.25, 0.5, 1 ,2, 4, 8 kHz PT averages for low, medium and high frequency groups	Illumina 370 K, Affymetrix 500 K, Illumina 700 K, Illumina Hap550, HumanExome Chip / 8,455,987* (INGI-FVG), 6,490,547* (INGI-CARL), 38,020,975* (TALANA), 6,302,431* (CILENTO), 6,664,949* (SR).	N = 2636 (meta-analysis)	European, Silk Road	*PCDH20, SLC28A3*
Fransen *et al.*, 2015[38]	PC scores from: Air conduction 0.25,0.5,1,2,3,4,6,8 kHz Bone conduction 0.5,1,2,4 kHz.	Illumina CNV370 quad chip, Illumina HumanOmniExpress BeadChip / 11,626,570* (before QC).	N = 2161 (quantitative)	Belgian	-
Hoffman *et al.*, 2016[113]	Electronic health records – ICD coding Self-report measures and SRT scores	Affymetrix Axiom arrays (optimised for population subgroups) / 9,469,183*(EUR), 8,090,486*(LAT), 6,517,021*(EAS), 7,829,026*(AFR)	N = 6527 cases, N = 45,882 controls	White, non-Hispanic American	*ISG20, TRIOBP*

Table 1.5. Summary table of ARHI GWAS samples and phenotype methods to date.

Study, first author and publication year; Phenotype, the phenotype used for association analysis; Genotyping array / no. of SNPs; genotyping array used to genotype discovery sample and the number of SNPs included in discovery analysis. *denotes the number of SNPs following imputation. [a] denotes number of SNPs prior to QC in the 1958 British birth cohort sample; Ancestry of samples, ancestry of sample used in discovery analysis; Genes identified P<5E-08, genes that contain SNPs or are in close proximity to SNPs that reached genome-wide significance in association analysis.

A study published in 2016 used self-report and SRT data obtained from electronic health records. This permitted a high powered GWAS of >6,500 ARHI cases and >45,000 controls, where association was identified with two novel SNPs at genome wide significance. This discovery sample comprised of a subset of the GERA cohort (included only non-Hispanic whites). The most highly associated SNPs were further investigated in the remaining ethnic groups within the sample, and under different ARHI phenotype classifications (SRT and speech discrimination score, SDS and self-report measures). Two novel SNPs were identified in *ISG20* and *TRIOBP*, and the associations were replicated. The authors also then took a candidate gene approach and identified two novel SNPs in *ILDR1* and *EYA4*[113]. Prior to this study, no previous ARHI GWAS had identified and replicated multiple genome-wide significant risk variants. Table 1.5 provides an overview of all previous ARHI GWAS study samples and phenotype classifications.

A limitation of current GWAS published on a broad range of traits is that findings only explain <20-30% of the trait variance[220]. ARHI is thus far no exception to the missing heritability problem, in that the current genetic variants identified via GWAS largely do not account for trait heritability estimates. For complex traits, this is thought to be in part due to the way heritability is measured; inconsistencies in methodologies, the samples used and inflated measures due to most methodologies taking into account only additive genetic components resulting in 'phantom heritability'[220], a term used to describe the disparity between heritability estimates and the effect sizes of known pathogenic variants. Epigenetic factors are also likely to account in part for the missing heritability of complex traits[221] such as ARHI[222]. Some epigenetic marks are inherited while some are acquired and are more dynamic. They play a role in all traits, predominantly in gene regulation, an important factor to study in ageing traits.

In terms of ARHI, there have been a limited number of significant associations identified, and all with relatively small effect sizes. Therefore, it may be that the current 'missing heritability' in ARHI is due to a combination of phantom heritability, a number of rare SNPs with larger effect sizes which are not currently on GWAS arrays, or common SNPs with small effect sizes which can only be identified with much larger sample sizes than previous ARHI GWAS have used. The latter is likely under the common disease, common variants hypothesis. With the advent of large-scale datasets with which to perform GWAS, this hypothesis that ARHI genetic risk comprises of multiple common SNPs with small effect sizes, can now be tested (as detailed below), and forms the basis of the work in this thesis.

The UK Biobank cohort

Over the past five years there has been a marked change in the size of cohorts that are available for GWAS analysis. Previously, substantial sample sizes were only attenable via meta-analyses, but biobank-style cohorts are now being created, which is collecting data on hundreds of thousands of participants. The UKBB was a leader in this movement and is the main dataset studied in this thesis. The UKBB was created as an international health resource, comprising of data on over 500,000 volunteers living in the UK and aged between 40-69 at the time of recruitment in 2009. Volunteers were invited to participate in the UKBB based on their postcode proximity to an assessment centre within the UK, and their age. This means that for all measured traits, there was an unbiased recruitment strategy; something which is difficult and not practical to achieve in a more traditional research recruitment strategy.

The UKBB resource is being extensively used for epidemiological and genetic research internationally; over 700 papers that used data from the resource, have been published. A large proportion of these publications are GWAS studies, including a number on sensory disorders, cognitive function and ageing traits[223–225]. UKBB is the first and largest multi-trait resource with which high powered GWAS can be performed. The resource therefore provides an opportunity for epidemiological and genetic research into multiple complex traits such as ARHI.

The size of the UKBB means that the statistical power of genetic analysis is elevated such that SNP associations with polygenic complex traits can be identified. As the main limitation of ARHI GWAS prior to this work was the limited statistical power, alongside the hypothesis that ARHI is highly polygenic, this dataset creates a unique opportunity to conduct the first substantially powered ARHI GWAS. During the 'GWAS era' multiple methods have been devised to perform association analysis. Analytical methods have been adapted for the large-scale datasets as they present new challenges. For example, due to its size of over 500,000 samples, the UK Biobank dataset poses new challenges such as computational efficiency and within-sample population structure. Previous GWAS had not featured samples of this magnitude without taking a meta-analysis approach.

The methods are continually evolving and did so rapidly while the work outlined in this thesis was undertaken. The development of such methods is determined in-part by the type of data that is available. In the case of this study, the following parameters were considered when identifying a suitable method to use for association analysis:

phenotype data format (binary or quantitative), computational efficiency and practicality, and accountability for population structure and relatedness (effects that are commonly present in datasets of a substantial size[226]).

At the time of analysis, PLINK 2.0[227], BOLT-LMM[228], SAIGE[229] and SNPTEST[230] were all available and considered for use. While all four methods could accommodate a binary phenotype, only BOLT-LMM and SAIGE could account for within-sample relatedness. BOLT-LMM was favourable as the relationship matrix was computed in conjunction with the association analysis. SAIGE, however, required this step to be conducted prior to association analysis. At the time of analysis, this was computationally demanding and was not a practical approach with the means available. In addition, PLINK 2.0 and BOLT-LMM had the quickest analysis speed. BOLT-LMM was therefore selected as it was suitable for the binary phenotype, accounted for within-sample-relatedness, was computationally efficient and relatively fast. PLINK 2.0 and GEMMA were used for association analysis on the replication samples, due to sample-specific constraints, as described in Methods 2.1.

Post-GWAS analysis

Since the advent of GWAS, much progress has been made in discovering genetic variants that are associated with complex traits. However, much of the true progress from GWAS findings will be in the accurate diagnosis, treatment and prevention of such conditions. This is only possible if GWAS associations are translated into biological and clinically relevant information. The aim of identifying these genetic variants is ultimately to discover which pathways and mechanisms are involved in a pathology. For many conditions, these pathways and mechanisms are more viable targets for diagnosis, prevention and treatments than single SNP variants. Post-GWAS analysis is therefore a crucial stage in a GWAS study protocol.

Post-GWAS analysis is a multi-step process tailored for the individual study, based on the number and nature of SNP associations, a priori work on the gene variants and the specific trait of interest. Generally, the first step is to validate the associations in order to eliminate false positive results and study-specific effects. Most commonly, a replication of the discovery association analysis is sought in an independent sample. Where possible, tests are performed on the same genetic markers under the same analytical methods and with a phenotype definition consistent with that used in the discovery analysis[231].

Moreover, the causal SNP at any one locus may not have been included in the GWAS analysis as only a subset of SNPs in the genome feature on genotyping arrays and

imputation reference panels. SNPs in close proximity to the causal SNP (whether it was included in the analysis or not) will have inflated test statistics due to LD between the SNPs. Identifying the number of causal SNPs at a locus can therefore be a challenge. Methods have been devised that can predict the number of causal SNPs at reach region of association, and which SNP best captures the variance. The estimate also predicts whether the test statistics for this SNP likely either inflates or diminishes the true effect size for a region[232]. For example, if multiple SNPs at a locus influence a trait but act in opposite directions, the reported effect size is therefore reduced when only one SNP is used to define the association within an LD region. Further methods can then be used to identify the most likely causal SNP.

As single SNP associations alone do not yield a great amount of biologically relevant information, the combined effects of multiple SNPs can be studied via gene-based or gene-set analysis. Here, genes in close proximity to lead SNPs (identified in association analysis) are grouped based on either genomic position, known biological functions, states and processes or by expression signatures. The analysis identifies pathways or biological processes that show enrichment for these gene sets. Hypotheses can then be devised, regarding links between these pathways and processes, and the trait of interest. Furthermore, identifying biological mechanisms and pathologies common to multiple complex traits can advance the understanding of each trait individually, and the interactions between traits.

There are several tools that can be used for *in silico* functional annotation. Those that perform gene-set analysis and pathway analysis such as FUMA[233] (containing MAGMA[234]), DEPICT[235], DAVID[236] and GARFIELD[237], are commonly used following association analysis. Selecting which tool to use can be based on several criteria, including: (i) the type of input data; either summary statistics from an association study or a pre-defined gene list, (ii) the reference datasets that the tool uses; some datasets are more comprehensive than others and can include tissue-specific expression data that is relevant to the trait of interest, and (iii) what the output will be used for; whether comparisons will be made with complimentary data or if independent conclusions will be drawn. A discussion of a suitable tool to use for this study is presented in section 2.3.1.

Following *in silico* functional analysis, it is common to employ *in vitro* and *in vivo* methods to investigate the biological function of genetic risk loci on the tissue or organ of interest. In terms of ARHI, the human auditory system is anatomically difficult to obtain, including post-mortem, as it is encased by the temporal bone. Therefore, in order to conduct histological and molecular work, a suitable animal model is required with which to study

auditory function, damage and dysregulation. Furthermore, the use of animal models permits *in vivo* analysis and genetic modification techniques.

The mouse is the model organism of choice for ARHI genetics. The mouse auditory system has a similar physiology and structure to that of humans[238] and protocols for assessing hearing function in mice are well established, such as auditory brain response (ABR) testing[239] or the less informative startle response[239]. Furthermore, a substantial proportion of genetic information is conserved between the two species; humans and mice share 85% of their protein-coding genome. While there is a substantial divergence in functions such as chromatin organisation and transcriptional regulation[221], an extensive toolkit is available for the genetic manipulation of mice[239], which is not possible in humans. This includes but is not limited to, conditional knockouts, knock-ins, insertion of reporter transgenes and CRISPR transgenics (this relatively new gene editing technique has already been used in auditory research, such as for preventing deafness in *Beethoven* mice by disrupting the mutant allele of the *Tmc1* gene)[240].

ARHI genetic studies with mice predominantly comprise three broad methods[239]: (1) the identification of pathogenic variants in mouse strains that have a genetic predisposition to express a hearing phenotype (2) large-scale mutagenesis screens, and assess the progeny for hearing phenotypes with the aim to identify potential HL risk loci. and (3) engineer mouse strains to express mutations in candidate HL genes and evaluate the resulting pathology. ARHI candidate genes that have been tested in human populations were first identified in mice, via methods (1) and (2).

Firstly, several mouse inbred strains carry genetic variants that predispose them to develop ARHI. Distinct risk variants in the different strains, result in distinct pathologies. For example, the C57BL/6J-Tyr$^{c\text{-}2J}$ strain does not produce strial melanin and has a reduced endocochlear potential, indicative of a role of reactive oxygen species in the pathology of ARHI[238]. The C57BL/6J strain (which carries the Cdh23^{ahl} allele) displays a progressive loss of inner and outer hair cells and the loss of spiral ganglion cells at the basal turn, by seven months of age [238,241]. This is the most commonly used strain in transgenic studies, and so the accelerated HL in knock-out models must be accounted for in HL study conception and the interpretation of any subsequent findings. In recent work, the gene defect in C57BL/6J mice has been rescued in a subset of the strain, by CRISPR/Cas9-mediated homology directed repair [238]. GWAS can also be conducted in animal models; a study identified nine significant loci in a GWAS conduced on 99 inbred strains of mice, with phenotypes classified by ABR thresholds[242]. Identifying the risk

variants in strains that are pre-disposed to developing deafness, or via mouse GWAS, gives rise to human ARHI risk gene candidates.

Secondly, mutagenesis screens have proved to be a successful alternative to identifying novel hearing loss genes. Most recently, a mutagenesis screen performed on 3006 knockout mouse strains identified 67 hearing gene candidates, 52 of which were novel[243]. In addition, findings from some mutagenesis screens have already been validated in human populations. In 2016, a group observed an association between the *stdf* SNP in the S1pr2 receptor and with a decrease in endocochlear potential and increased stria vascularis degeneration. The association was replicated in the 1958 British Birth Cohort sample via candidate gene analysis[244]. While this study has been successful in applying the finding to a human population, without having determined the extent and nature of the human ARHI genetic risk, it is difficult to establish the extent to which mice are a suitable model to study human ARHI genetic risk and thus how applicable these findings are.

Thirdly, for the reasons listed above, by using mice, there is an opportunity to engineer strains to display genetic variants identified in population GWAS or human candidate gene studies. The type of pathology which is then observed in the mouse can give rise to a hypothesis of the corresponding human pathology present in the GWAS population sample. However, we first need to identify risk variants in human populations as the genes identified prior to the work in this study are sparse, and account for a small proportion of the genetic variance of the trait.

1.5 Thesis aims

ARHI is the most common sensory impairment in the ageing population, and a major risk factor for dementia, yet little is known about the underlying biological determinants. Due to the location of the auditory system, encased in the temporal bone and as part of the brain, histological analysis is challenging. Genetic analysis is an alternative route by which to identify and investigate such biological determinants. However, prior to this study, little was known about the genetic component of ARHI, or whether genetic risk factors contribute to ARHI in the general population.

This is mainly because previous ARHI studies have lacked the necessary statistical power due to limited sample sizes and the absence of a comprehensive, consistent definition of ARHI. ARHI is a polygenic, heterogeneous trait and, under the common trait common variant hypothesis, large, highly powered studies are required to identify genetic variants underlying such traits. It is challenging to develop diagnostics, treatments or

prevention strategies for ARHI without a comprehensive understanding of the biological determinants of the trait.

The overall aim of this thesis is to develop current knowledge of ARHI pathogenesis by identifying genetic risk variants and further investigate putative mechanisms and pathways involved in ARHI. The UKBB resource contains phenotypic and genetic data on half a million volunteers from the UK, aged 40-69 and therefore provides a unique opportunity to investigate the genetic risk of multiple conditions, including ARHI, with unparalleled statistical power. With this motivation, I set out to address three specific aims:

1. Use the UKBB data resource to derive phenotypes that best describe age-related hearing loss in the population sample, and that are suitable for genetic association analysis

Previous ARHI GWAS phenotypes have predominantly been derived from PTA thresholds. However, such data (i) is not practical to collect the large scale that is required (ii) does not encompass a key symptom of ARHI; difficulty interpreting speech in the presence of background noise. Therefore, in the UKBB, alternative data fields are investigated in order to identify a measure that best describes ARHI and that is suitable to use in genetic association analysis.

2. Perform genetic association analysis with surrogate ARHI phenotypes derived from UKBB data fields, with the aim to identify putative ARHI risk loci.

The genetic risk factors of ARHI are largely unknown. By identifying associations with SNP variants, genetic risk factors can be identified, and their biological functions investigated. GWAS has yet to be proven as an adequate method to ARHI genetic risk loci. Although hypothesised to be a polygenic trait comprising multiple risk variants of small effect size, a population study is yet to prove this theory. The UKBB resource can be used to test the theory that ARHI genetic risk variants can be detected with sufficient statistical power, and that there are multiple such variants with small effect sizes.

3. Investigate the putative functions of genetic risk variants identified in association analysis, in the development of ARHI

The nature of ARHI risk loci is largely unknown; genes implicated in ARHI risk may be common to other forms of hearing loss, and variants may result in similar defects as seen in congenital HL but at a later onset, or, alternatively, ARHI may be caused by distinct genes and biological mechanisms, only observed in cases of ARHI. Investigating putative

functions of the identified risk loci will give rise to new hypotheses of how ARHI develops and which auditory structures are affected.

Chapter 2 - Methods

2.1 Population study samples used in Chapters 3 and 4

The UKBB is the primary population sample that is used in Chapters 3 and 4. Two further samples, TwinsUK and the English Longitudinal Study of Ageing, are used in Chapter 4 for GWAS replication analysis. The following sections contain details on data that is available on each sample, the recruitment method and cohort demographics.

2.1.1 UK Biobank

The UKBB is an international health resource which is following the health and wellbeing of over 500,000 volunteers who were recruited in the UK and were aged between 40-69 years during the time period 2006-2010. The Biobank was conceived to collect data with the aim to 'improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses' (https://www.ukbiobank.ac.uk/). Data collection consisted of a number of physical measurements, questionnaire data and biological samples including blood, saliva and urine[245]. A subset of the sample, consisting of 20,000 participants residing in north east England, were invited for a repeat visit in 2012-2013. The data collected by UKBB is available for use by researchers, subject to a successful application. This work presented in this thesis is associated with the UKBB project numbered 11516.

2.1.1.1 UKBB non-genetic data

The phenotype measures collected by UKBB that are specifically relevant to this thesis are explored in detail in Chapter 3. These consist of a speech in noise hearing test and three hearing-related questionnaire measures that were included in the health and lifestyle questionnaire. In Chapter 3, UKBB data is assessed for use as a surrogate phenotype measure for ARHI in the population sample, to be used for genetic association analysis. A description of the approaches used to study the UKBB speech in noise test and hearing- related questionnaire measures are described alongside the results presented in Chapter 3. The work in Chapter 3 is predominantly conducted with R versions R 3.2.0 – R 3.6.1.

Table 2.1 displays the description, instance and Field ID for all non-genetic data fields used in this work. For example, where the speech in noise (SIN) test baseline data is studied in Chapter 3, (section 3.2) the following two data fields are used: 'Mean signal-to-noise ratio (SNR), (left)' Field ID 4230 and 'Mean signal-to-noise ratio (SNR), (right)' Field ID 4233. In Chapter 3 these are referred to as 'SRT_L' and 'SRT_R' respectively. In addition to the baseline data collection, a subset of participants competed the SIN test at the Cheadle assessment centre. For all longitudinal or repeat SIN data, the following additional data fields are used: 'Mean signal-to-noise ratio (SNR), (left)' at the first repeat assessment visit (2012-2013) Field ID 4230.1.0 and 'Mean signal-to-noise ratio (SNR), (right)' at the first repeat assessment visit Field ID 4233.1.0.

Description	Instance	Field ID
Speech-reception-threshold (SRT) estimate (Left)	T0	20019.0.0
Speech-reception-threshold (SRT) estimate (Right)	T0	20021.0.0
Speech-reception-threshold (SRT) estimate (Left)	T1	20019.1.0
Speech-reception-threshold (SRT) estimate (Right)	T1	20021.1.0
Hearing difficulty/problems	T0	2247.0.0
Hearing difficulty/problems	T1	2247.1.0
Hearing difficulty/problems with background noise	T0	2257.0.0
Hearing difficulty/problems with background noise	T1	2257.1.0
Hearing aid user	T0	3393.0.0
Hearing aid user	T1	3393.1.0
Age when attended assessment centre	T0	21003.0.0
Age when attended assessment centre	T1	21003.1.0
Sex	T0	31.0.0
Date of attending assessment centre	T0	53.0.0
Date of attending assessment centre	T1	53.1.0
UKBB assessment centre	T0	54.0.0
UKBB assessment centre	T1	54.1.0

Table 2.1. UK Biobank data fields and data field IDs.

Description, description of data field, as on UKBB data showcase; Instance, instance of assessment, 'T0' indicates baseline assessment and 'T1' indicates the first repeat assessment visit (2012-2013); Field ID, UKBB unique field ID.

2.1.1.2 UKBB phenotype definitions

The aim of Chapter 3 is to assess various hearing-related data fields in the UKBB data resource and establish whether they sufficiently measure ARHI in the cohort, and whether they are suitable measures to use as phenotypes for GWAS. Therefore, much of

the methodology of how the two phenotypes describe hearing difficulty (*HDiff)* and frequent hearing aid use (*HAid*) were derived, is presented alongside the results in in Chapter 3 and at the beginning of Chapter 4. The *HDiff* and *HAid* phenotype definitions are the same for the UKBB discovery and UKBB replication samples. Participants are assigned case/control status based on their responses to questionnaire measures regarding hearing difficulty and hearing aid use. *HDiff* cases responded "Yes" to both questions "Do you have any difficulty with your hearing" (UKBB Field ID 2247) and "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" (UKBB Field ID 2257) *HDiff* controls were selected if their response to both questions was "No". Participants with any other combination of responses are removed. In addition, *HDiff* controls aged <50 are removed from analysis, as were any controls that responded "Yes" to the question "Do you use a hearing aid most of the time?"

HAid cases responded "Yes" to "Do you use a hearing aid most of the time?" (UKBB Field ID 3393) and controls responded "No". If participants answered the questionnaire twice, i.e. attended an assessment centre for a repeat visit, the answer at the second time point was used in analysis, in order to increase the mean age of the sample. To reduce the likelihood of including congenital forms of deafness, participants who selected 'I am completely deaf' in the UKBB questionnaire were excluded from analysis.

2.1.1.3 Study sample used for genetic association analysis

The majority of UKBB participants are of self-reported white ethnicity (94.6%). As the largest population group, the northern European sample was therefore used in Chapter 4 for genetic association analysis as this would provide the greatest statistical power. The sample used in the discovery association analysis was selected based on having 'white British' ancestry, a classification derived from both principle component (PC) analysis and self-declared ethnicity[246]. The subset of the UKBB sample that is used in the replication meta-analysis is the remaining subset of samples with northern European ancestry.

Maxim Freidin kindly assigned participants into discrete ancestry clusters in the UKBB, using the 1st and 2nd PC vectors provided by UKBB. A k-means clustering algorithm was applied to generate clusters for each PC. The cluster indices were combined for the PCs (1.1, 1.2, …, 5.5), and compared against self-reported ancestry and ancestry groups assigned accordingly. If contradictory, the pairwise clusters take precedence over the self-report grouping.

2.1.1.4 UKBB genotyping and imputation

The ~500,000 samples in UKBB were genotyped on one of two arrays; 50,000 samples were genotyped on the Affymetrix UK BiLEVE Axiom array while the remaining ~450,000 were genotyped on the Affymetrix UK Biobank Axiom® array. The two arrays shared 95% coverage resulting in >800,000 genotyped SNPs. Imputation was carried out centrally by UKBB, primarily using the HRC reference panel and IMPUTE2[247]. SNPs which do not feature on this panel were imputed with the UK 10K and 1000G panel. Analysis in this study was conducted with version 3 of the UKBB imputed data with 487,409 samples imputed and available for analysis following UKBB centrally performed SNP and sample quality control (QC).

In addition to QC performed centrally by UKBB, for this study samples with excess heterozygosity, excess relatedness and sex discrepancies were identified (using a QC file centrally provided by UKBB) and removed prior to association analysis. Where samples show excess heterozygosity and sex discrepancies (with the participant's self-report status), samples are removed as these can signify sample contamination or incorrect sample IDs. 'Excess relatedness' in this context refers to samples which have more than 10 putative third-degree relatives in the kinship table. Such samples were removed to reduce effects from population stratification within the sample.

2.1.1.5 Genetic association analysis with the UKBB sample

For both the discovery and replication association analysis with the UKBB sample subsets, a linear mixed-effects model approach to test for association between imputed SNP dosages and the two traits. BOLT-LMM v.2.3.2[226] was used for the association analysis, and corrects for population stratification and family structure mean that it is an optimal program to use for a dataset of this size.

The analysis was also adjusted for age, sex, UKBB genotyping platform and UKBB PCs1-10. For QC SNPs were filtered based on two thresholds: (1) minor allele frequency (MAF) > 0.01; and (2) imputation INFO score > 0.7. By implementing an MAF cut-off of 0.01, the likelihood of including participants with forms of congenital deafness was reduced; only variants that occur at least in 1/100 participants were detected, which is a higher frequency of variants than the frequency of congenital deafness. Individuals with < 98% genotype call rate were removed. UKBB PCs 1-10 were included as covariates to further reduce effects of population stratification and underlying population or sample structure within the dataset. Likewise, the UKBB genotyping platform was included as a covariate to reduce the effect of platform-specific genotype calls. Age and sex were included as

covariates due to the relationships between these factors and the two phenotypes, as explored in Chapter 3. A standard threshold was implemented for SNP INFO score, a measure that resembles the imputation certainty of an individual SNP; an INFO score of 1 signifies complete certainty, while an INFO score of 0 signifies complete uncertainty.

2.1.2 TwinsUK

The TwinsUK cohort is used for replication analysis in Chapter 4. TwinsUK is the largest adult twin registry in the UK and comprises over 13,000 healthy twin volunteers aged 18-103. The cohort is predominantly female (>80%) with a balance of monozygotic and dizygotic twin pairs[248]. Collection of data and biological materials commenced in 1992 and is currently ongoing. During study participation, twins regularly complete health and lifestyle questionnaires and attend the study centre for clinical evaluation.

2.1.2.1 TwinsUK study sample

TwinsUK was selected as a suitable cohort to use for replication analysis as the study volunteers predominantly have Northern European ancestry and data collection included relevant questionnaire data on hearing status. From the main cohort, participants that were included in the *HDiff* study were those that had responded to the question 'Do you suffer from hearing loss?' in one of the TwinsUK study questionnaires. Participants that were included in the *HAid* analysis, were those that had responded to the question 'Do you wear a hearing aid?' or have confirmed 'Wearing a hearing aid'. Only twins aged >40 were included in the analysis, to be comparable to the UKBB sample.

2.1.2.2 TwinsUK phenotype definition

The TwinsUK phenotypes were likewise derived from responses to questions. *HDiff* cases responded either "Yes, diagnosed by doctor or health professional" or "Yes, not diagnosed by health professional" to the question "Do you suffer from hearing loss?" while controls responded "No". *HAid* cases responded or indicated "Yes" to either of "Do you wear a hearing aid?" and 'Wearing a hearing aid'. *HAid* controls responded "No".

As TwinsUK is a longitudinal study, several participants gave responses to the same questions on multiple occasions. The most recent response was included in analysis, unless the latest response indicated that hearing had improved. In this scenario, the participant was excluded. If the individual's response to any of the questions was the same as the previous response or indicates hearing has worsened, the second response and response date was taken. If response indicated hearing has improved, the sample was removed.

2.1.2.3 TwinsUK genotyping and Imputation

Genotyping of TwinsUK was conducted with the following combination of Illumina arrays: HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M. The imputation reference was 1000G Phase3 v5 (GRCh37).

2.1.2.4 Genetic association analysis with the TwinsUK sample

Due to the nature and size of the TwinsUK cohort, the association analysis was conducted using a linear mixed-effects model regression adjusting for age and sex with GEMMA[249]. GEMMA is a tool that can powerfully account for family structure present in cohorts such as TwinsUK. Maxim Freidin kindly ran the association analysis on the TwinsUK cohort with the phenotypes that I derived from the TwinsUK questionnaires.

2.1.3 English Longitudinal Study of Ageing

2.1.3.1 ELSA study sample

Data from the ELSA cohort was utilised in the replication analysis presented in Chapter 4. Similar to TwinsUK, ELSA was selected as a suitable cohort to use for replication analysis as the study volunteers predominantly have Northern European ancestry and data collection included relevant questionnaire data on hearing status. ELSA is a longitudinal study, consisting of around 12,000 respondents from the Health Survey for England[250]. The study began recruitment in 2002 and participants have been followed-up every two years. Eight waves of data collection have been completed since 2002. The sample consists of men and women residing in the UK and that are ≥50 years of age.

2.1.3.2 ELSA phenotype definition

For the ELSA sample, phenotypes were derived from responses to questionnaire measures collected during study Wave 7. This wave was selected as the questions best matched those collected on the UKBB sample, and it provides the largest sample with a consistent questionnaire and with genotype data.

The ELSA *HDiff* phenotype was derived using responses from two questions; "Do you ever have any difficulties with your hearing?" and "Do you find it difficult to follow a conversation if there is background noise, such as TV, radio or children playing (using a hearing aid as usual)?" Cases were defined as participants who responded "Yes" to both questions, and controls who responded "No" to both questions. As in the UKBB analysis, controls who reported hearing aid use or age <50 were removed, as were any cochlear implant users in the case or control samples. The *HAid* phenotype was derived using the question "Whether ever wears a hearing aid"; cases responded "Yes most of the time", or

"Yes some of the time" while controls responded "No". During ELSA data processing, age was capped at 90 years, and thus individuals aged > 90 were reported to be 90 years of age.

2.1.3.3 ELSA genotyping and Imputation

ELSA samples were genotyped at UCL Genomics in two batches using the Illumina HumanOmni 2.5M platform. Imputation was carried out centrally by ELSA with IMPUTE2, using the 1000 Genomes phase I data set (https://www.elsaproject.ac.uk/uploads/elsa/elsa_analysis.pdf). Prior to association analysis, centrally at ELSA before data release, sample relatedness was estimated in PLINK 1.934 and subsequently one of each pair of related individuals was excluded.

2.1.3.4 Genetic association analysis with the ELSA sample

The ELSA sample sizes for *HDiff* and *HAid* analysis are <5,000. BOLT-LMM authors advise alternative methods for samples that are n<5,000 due to BOLT-LMM being unreliable when used on samples of this size (https://data.broadinstitute.org/alkesgroup/BOLT-LMM/). Therefore, PLINK2 was used to carry out logistic regression to test for association in the ELSA sample, adjusting for age and sex, as in the UKBB analysis.

2.2 Post-GWAS statistical analysis

2.2.1 Linkage Disequilibrium Score Regression

A common quality control step post-GWAS is to calculate the genomic inflation factor (λ_{GC}) which estimates the inflation of the test statistics. Inflation can be a sign of confounding in the analysis, caused by factors such as cryptic relatedness or population stratification present in the sample. However, in more recent GWAS that use large sample sizes to study polygenic traits, the λ_{GC} cannot distinguish between the inflation caused by these confounding factors and inflation that is caused by polygenic effects. In these cases, alternative methods can be used to estimate the proportion of the observed inflation in the χ^2 statistic that the intercept attributes to factors other than polygenic effects, such as confounding from cryptic relatedness. In Chapter 4, the univariate linkage disequilibrium score regression (LDSC) method is applied for this purpose, which regresses SNP LD scores against the SNP test statistics. The online tool used for this analysis is described in detail elsewhere[251,252].

2.2.2 Conditional and Joint analysis

As noted above to linkage disequilibrium (LD), SNPs that are in LD with a pathogenic variant will have inflated summary statistics. This means that it can be difficult to determine whether there are secondary associations at a region of association, or whether the test statistics of SNPs within a region represent one signal and are inflated because they are in LD with the functional variant. Methods have been devised which can determine whether there are secondary associations at a locus, and whether the lead SNP (SNP with the lowest P-value in the association analysis) captures the effect of the association at the locus. Conditional and joint SNP analysis was used in Chapter 4 to identify independent signals within the highly associated regions, using GCTA-COJO[232]. This analysis requires the LD reference sample, which was obtained by random selection of 10,000 individuals from the UKBB cohort with White British ancestry. The reference sample size of n=10,000 was selected to maximise power based on previous data simulations[232].

2.2.3 Heritability calculation

SNP heritability (h^2_g) was estimated independently for the *HDiff* and *HAid* traits. The estimate was calculated with BOLT-LMM using restricted maximum likelihood variance analysis. As the two traits are not quantitative measures, the estimates were recalculated to the liability scale. The prevalence is required in order to re-calculate the estimate. The trait prevalence specified in this calculation was the case prevalence in the analysed sample, and therefore was 0.35 for *HDiff* and 0.052 for *HAid*.

2.2.4 Replication association analysis

Independent replication of the association was conducted via a meta-analysis of the non-British White UKBB subsample, the TwinsUK sample and the ELSA sample. A meta-analysis was conducted to increase the statistical power of the replication analysis by increasing the total sample size.

2.2.4.1 Association analysis and meta-analysis

Due to the different nature of the three samples, three different methods were used to conduct association analyses, as discussed in the section above. Firstly, PLINK2.0 logistic regression was used to test for association between the HDiff and HAid phenotypes and genetic data in the ELSA sample. PLINK2.0 was selected as the samples are n<5,000 and therefore not suitable for analysis under the BOLT-LMM method. Secondly, association analysis on the TwinsUK HDiff and HAid samples was performed with a linear mixed-

effects model regression with GEMMA. This method is suitable for the relatively small size of the sample, and the relatedness present within the sample. Lastly, BOLT-LMM was used to test for association with the candidate loci in the UKBB white non-British replication sample, as described for the discovery association analysis.

The lead SNPs identified with conditional analysis on the UKBB discovery sample association analysis were tested for association with *HDiff* and *HAid* phenotypes in a fixed-effect inverse-variance weighted meta-analysis using METAL[253] version 2011-03-25, with the White non-British UKBB, ELSA and TwinsUK samples. BOLT-LMM does not report analysed sample size per SNP, so to obtain the weight of the UKBB replication sample per SNP, the sample size reported is from the sample sizes from PLINK linear regression. Bonferroni significance thresholds were calculated for the two traits as 0.05/41, p=0.0012 for *HDiff* and 0.05/7, p=0.00714 for *HAid* based on the number of statistically significant independent associations in the discovery.

2.2.4.2 Replication study power calculations

A power calculation was performed to estimate the power of the meta-analysis replication for each independent locus analysed in the replication analysis at nominal significance ($p<0.05$) for both traits and at Bonferroni-corrected significance. ($p<0.0012$ for *HDiff* and $p<0.00714$ for *HAid*).

The following R code was applied:

```
chi<-qchisq(1-alpha/nl,1) # χ2 for significance threshold for the number of loci to replicate (nl)

t2<-(b/se)^2 # test-statistic for a SNP in discovery; b and se are effect size and corresponding standard error

q2<-t2/nd # proportion of variance explained by a SNP; nd is sample size in discovery

ncp<-nr*q2/(1-q2) # non-centrality parameter; nr is sample size in replication

pchisq(chi,1,ncp=ncp,lower.tail = F) # power of replication
```

nl=1 was used to estimate the power to replicate at nominal significance; nl=7 and nl=41 were used to estimate power to replicate SNPs at a Bonferroni-corrected significance threshold for *HAid* and *HDiff* respectively.

2.3 *In silico* functional analysis

2.3.1 Gene-set analysis

Gene-set analysis was one of the post-GWAS methods used in this work. The results are presented in Chapter 4. The goal of gene-set analysis is to provide functional context for multiple genetic associations and can be performed using multiple methods and datasets. In this work, gene-set analysis was performed via the online tool FUMA[233]. There are two distinct parts to the analysis: SNP2GENE and GENE2FUNC. SNP2GENE assigns lead SNPs based on the LD structures of associated regions and assigns lead SNPs to genes within a distance of 10kb based on positional mapping, eQTL information and 3D chromatin interactions. The degree of association is calculated for each gene within 10kb of a lead SNP.

GENE2FUNC (gene enrichment analysis) is then performed by testing whether any genes are overrepresented in specified pre-defined gene sets. The gene sets and gene ontology (GO) terms that are used in this analysis to classify the gene sets are from MSigDB v5.2 (http://software.broadinstitute.org/gsea/index.jsp). GO annotations include three groups: molecular function, cellular component and biological process.

FUMA[233] was selected as an appropriate tool as the two-step approach permits an input of summary statistics from association analysis, rather than a pre-defined gene list. For gene-set analysis, it deploys MAGMA[234], which has been shown to be statistically robust by comparison to alternative methods[254] and draws on the comprehensive MSigDB database (as noted above, v5.2 was used for these analyses). As of yet, there are no tools that can perform this analysis on data derived from cochlear tissue samples, as would be relevant for this study. Further, the same tool was used by colleagues and collaborators who were studying the shared genetic risk of Alzheimer's and ARHI. Obtaining a comparable output was advantageous when comparing enriched gene sets and for identifying common biological processes.

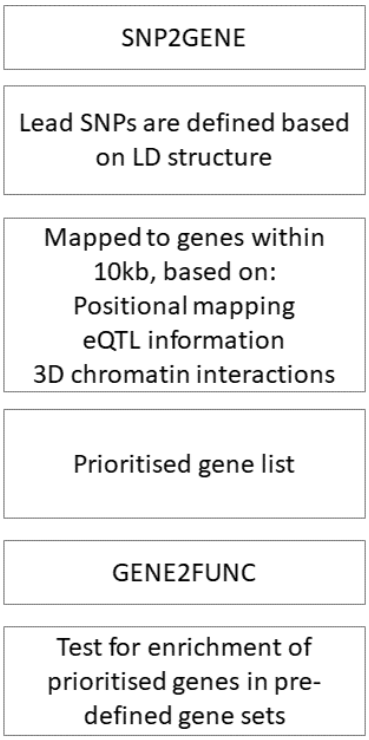

Figure 2.1. Flow diagram of the FUMA method for gene-set analysis

2.3.2 Genetic correlation analysis

Genetic correlation analysis was performed between the *HDiff* trait and 765 traits that have publicly available data for correlation analysis on LD hub[251,255]. Two traits that used similar phenotypes to that used in this study were removed; the summary statistics contributed by the Neale lab which utilised two self-report hearing traits from UKBB data. Summary statistics were only included in the input to LD hub where the SNPs were included in the recommended list (SNPs imputed to Hapmap 3 and removal of the MHC region on CHR6 26MB-34MB). A conservative significance threshold was calculated with a multiple-test correction (0.05/764, p = 6.5E-5). Due to the large number of significant findings, the results were further filtered based on the correlation size (rg). Traits with an rg >0.3 or < -0.3 were categorised into five groups based on trait symptom similarities: (1) breathing difficulties (2) health report/subjective wellbeing (3) hearing (4) low mood/depression and (5) pain.

2.4 Prioritisation of gene candidates for functional follow-up

Multiple methods were used in order to prioritise gene candidates for functional follow up analysis presented in Chapter 5.

2.4.1 Visualisation of genome regions of interest and LD calculations

Locus plots generated with LocusZoom[256] are useful for visualising SNP associations in the context of the genomic region at each lead SNP. It ought to be noted that a reference panel from the UK Biobank population (used in the genome-wide association analysis) is not available for this tool. It instead uses LD reference data from the hg/1000 Genomes Nov 2014 which is from the sample population group (EUR) and genome build (hg19). Similarly, LDlink[257] which was used to determine the LD between significant SNPs at loci in close proximity to the lead SNP, LDlink tool uses 1000 Genomes Project as reference. Here the British in England and Scotland (GBR) population subset from the 1000G project data were selected.

2.4.2 Predicating effect of variants on protein function using in-silico analysis

In chapters 4 and 5, VEP was used to interpret the results of the joint and conditional analysis. For chapter 4, lead SNPs were mapped to the nearest protein coding genes with a maximum distance of 10kb using VEP[258] GRCh37. In chapter 5, the genomic context of the lead SNPs was determined using VEP; whether the SNP was in a gene transcript, the presence and distance from gene transcripts and the predicted consequence of the variant. The predicted pathogenicity of missense variants was calculated with the Rare Exome Variant Ensemble Learner[259].

2.4.3 Data resources used in chapter 5 to gather data relating to primary gene candidates

In chapter 5, the criteria used to define a 'known' hearing gene that had a previous link to hearing impairment at the time of analysis was evidence of a link with the PubMed search terms 'gene name, hearing' and 'gene name, deafness' or when cross-referenced with the list of genes reported to underlie deafness in either humans or mice (2018)[260]. Genes were cross-referenced with the Jackson laboratory to determine availability of mouse models and hearing phenotype observations (https://www.jax.org/).

The gEAR database (https://umgear.org/) was also utilised to determine the expression level of candidate genes in mouse cochlear tissue. Data from an RNA-seq screen of cochlear hair cells and non-sensory cells in P0 mice was collected and used as evidence for gene prioritisation in chapter 5[261]. The data is presented in reads per kilo base per million mapped reads (RPKM).

2.5 Protein localisation analysis

Adult mouse cochlear tissue samples were prepared and stained with antibodies against genes of interest.

2.5.1 C57BL/6J mice

Mouse cochlear explants were obtained from C57BL/6J mice for use in Immunohistochemistry. The results are presented in Chapter 5. The C57BL/6J mouse strain are genetically predisposed to develop sensorineural HL. Mice were euthanised at P28-P30 in order to study the location of protein targets in an adult, fully developed auditory system, but before strain-specific deterioration of auditory structures are observed. The *Cdh23*^{ahl} allele in the C57BL/6J strain causes an accelerated HL, with a progressive loss of inner and outer HCs and loss of the SG cells at the basal turn of the cochlea by seven months of age[238,241], which progresses to a profound hearing loss or deafness at 18-24 months of age.

Sensory IHC loss and OHC loss is observed at the base of the cochlea in C57BL/6J mice as early as 3 months of age. As the degeneration progresses, OHC loss is observed throughout the cochlea at 26 months. At this age, a complete loss of IHCs can be observed at the base of the cochlea, while a 20% loss of IHCs is observed at the apex of the cochlea[262]. For these reasons, the mice used in this study were aged ~1 month (P28-P30). P28-P30 was selected because if using C57BL/6J mice at an older age, the loss of HC and SG cells would be occurring, and thus gene expression in these structures cannot be observed.

2.5.2 Mouse cochlear tissue preparation

The C57BL/6J mice used in this work were bred in an in-house facility. Mice were euthanised according to Schedule 1 procedures as described in United Kingdom legislation outlined in the Animals (Scientific Procedures) Act 1986.

The dissected inner ears were fixed in 4% paraformaldehyde diluted in phosphate buffered saline (PBS) for 1 hour at room temperature before being washed several times in PBS (3x 10 minutes at room temperature). The inner ear samples were then decalcified in 10% EDTA overnight at 4^{o}C before cochlear and vestibular system separation. Cochlea were mounted in 4% low-melting point agarose and sectioned on a Vibratome (1000 plus system, Intracel) at 200-µm intervals.

2.5.3 Immunofluorescence

The immunofluorescence protocol used in this work is summarised in Figure 2.2 and in the two sections below, 2.5.3.1 and 2.5.3.2.

2.5.3.1 Antibodies used to test for protein localisation

Antibodies used to identify protein localisation in the organ of Corti were: anti-nidogen-2 (NID2) at a 1:750 dilution (Ab14513, Abcam), anti-clarin-2 (CLRN2) at 1:1000 (HPA042407, Atlas Antibodies) and anti-rho guanine nucleotide exchange factor 28 (ARHGEF28) at 1:1000 (HPA037602, Atlas Antibodies). All were detected using of an isotype-specific Alexa Fluor 488 goat anti-rabbit secondary antibody (Santa Cruz Biotechnology). All antibodies were tested at various dilutions, in organ of Corti cell line number 2 (OC2) cells prior to tissue staining, in order to test and optimize the protocol.

2.5.3.2 Immunofluorescence and sample imaging

Antibodies were diluted in a goat blocking solution (4% triton, 8% goat serum, 1g BSA, 50ml PHEM buffer) and sections were stained with primary antibodies overnight at 4°C. Following PBS washes, sections were incubated with the secondary antibody at 1:1000 in darkness at room temperature for 2 hours. Phalloidin-Atto 647N to f-actin (Sigma-Aldrich, Gillingham, UK) and DAPI were added to the secondary antibody incubations at 1:1000 to stain hair cell stereocilia and DNA respectively. The samples were imaged using a Zeiss LSM 880 Airyscan 20x and 40x objectives.

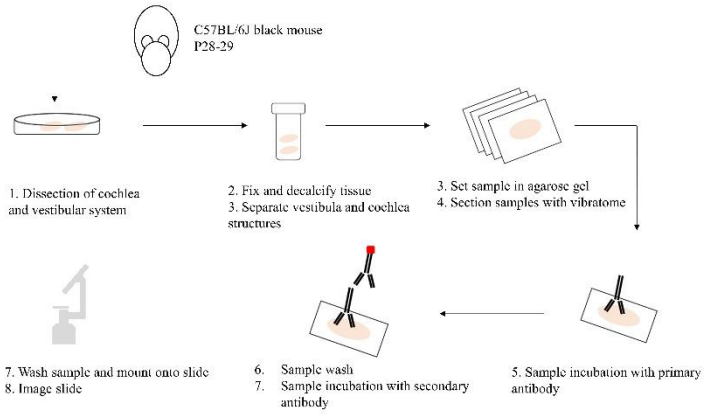

Figure 2.2. Diagram to summarise the immunohistochemistry method used in Chapter 5.

Chapter 3 - Defining a phenotype for association analysis

3.1 Introduction

As discussed in chapter 1, there is no standardised method with which to comprehensively classify ARHI in population studies. In order to conduct a valid GWAS however, the trait of interest ought to be unambiguously defined in the population. The aim of this chapter is therefore to identify a phenotype that best represents common, adult hearing impairment in the UKBB cohort and that is suitable to use in a GWAS.

Some traits are quantitative or qualitative by nature, while others are defined as such by the criteria used for measurement. Most traits or conditions that are quantitative are however classified into qualitative categories based on defined thresholds within the quantitative range. This classification aids diagnoses, treatments and epidemiological analysis. For example, BMI is a quantitative trait, but patients are subset into qualitative categories for clinical use (underweight, healthy, overweight, obese) based on pre-defined thresholds (<18.5, 18.5-24.9, 25-29.9, 30-39.9)[263].

In epidemiological studies, the prevalence of a condition is often reported, which requires the sample to be divided into 'affected' or 'unaffected' groups. In order to limit any bias that may be introduced by doing so, statistical definitions are commonly applied. A standard for defining a 'normal' (unaffected group) in a sample, is the selection of datapoints that are within 2 standard deviations from the sample mean[264]. However, this does not always reflect trait-specific clinical definitions such as symptom frequency. In these cases, arbitrary definitions of 'normal' are set, based on parameters from clinical, prognostic or operational definitions.

As far as possible, a definition for 'affected' cases in a population sample must be stable, unambiguous and precise. Likewise, bias must be minimalised in the definition and subsequent selection of 'unaffected' controls. To achieve this, where possible, controls are selected from the sample population sample as cases, will have been exposed to the same risk factors, and have had these factors measured in the same manner as the case sample[265].

As discussed in chapter 1, ARHI is a sensorineural, bilateral hearing loss that begins in the high frequencies and progresses with increasing age[266]. In the context of a GWAS on adult hearing loss, a phenotype would therefore depict a common adult hearing loss that progresses with increasing age. Results from PTA tests are commonly used where

available in population studies, as this is a clinical method used to diagnose a hearing impairment. The type of hearing loss that a patient presents with can be implied by the shape of the audiogram from the PTA test. A right-hand sloping audiogram for both ears signifies sensorineural, bilateral hearing loss; as expected in ARHI.

However, as discussed previously, while PTA testing is the standard and most widely used hearing test, it does not measure an individual's ability of speech perception (and thus possible hidden hearing loss), a common symptom of ARHI. PTA testing is also impractical to conduct on a large scale such as that necessary for genetic association analysis; the test requires specialist equipment, a quiet environment, a trained professional and can take up to thirty minutes. An alternative test, the speech in noise (SIN) test[187], can also be used to classify an individual's hearing status; it measures the subject's ability to correctly decipher speech in the presence of varying levels of background noise[28,185]. SIN testing therefore assesses a common symptom of ARHI (deciphering speech in the presence of background noise) and can be completed in the absence of a trained professional either online or on the telephone. Consequently, this test may be a more appropriate measure of ARHI in the general population and a more practical method for gathering phenotype data on the scale required for genetic association analysis.

Lastly, self-report questionnaire measures are increasingly used to define phenotypes for use in GWAS. Questionnaires permit large-scale data collection and can result in both quantitative and qualitative phenotype definitions depending on whether the question responses are graded or binary. Questionnaires can assess numerous details of a trait such as severity and persistence, which can be combined to derive phenotypes that capture details from more than one questionnaire measure alone. These are also not without limitations however such as the incorporation of subjective or emotional responses.

Of the previous eight GWAS conducted on ARHI, seven used phenotypes derived from PTA thresholds, as discussed in chapter 1. Individual studies have assessed hearing at different frequency ranges and used either a z-score or PC analysis to derive the phenotype for analysis[38,92,107,109,111,112]. Collectively, three loci were significantly associated with ARHI-related phenotypes in these studies. The most recent ARHI GWAS study used electronic health records along with self-report measures and SRT scores[113] rather than a PTA-derived phenotype. The size of this study was a magnitude larger than any previous ARHI GWAS and doubled the number of significant associations found in a single study; two significant ARHI-associated loci were identified and replicated. The use

of electronic health records and questionnaires to define the phenotype permitted the increase in sample size and thus increase in study power. The main limitation of ARHI GWAS to date has been the relatively small sample sizes used for analysis; globally there are few cohorts that have hearing data, genetic data and are of a sufficient size for a highly powered association analysis.

The work presented in this chapter makes use of data collected in the UKBB cohort[245] (summarised in - Methods, 2.1.1). Upon attendance at a UKBB assessment centre, participants donated biological samples for genotyping, completed lifestyle questionnaires, online tasks and had standard physical measurements taken. The collected data comprises multiple hearing-realted data fields, including those derived from the UKBB SIN hearing test which participants performed on a touch screen monitor while at an assessment centre. Secondly, hearing-related questionnaire measures were part of the health and lifestyle questionnaire that was also completed at the assessment centre on the touch screen monitors. In total, data were collected at 23 UKBB assessment centres across the UK, 21 of which were open during a period within the initial recruitment phase from 2007-2010. Cheadle and Newcastle assessment centres hosted the imaging and repeat assessments respectively (2012-13), see - Methods, 2.1.1.

Currently, no single measure can encompass whether an individual's hearing impairment is a sensorineural, bilateral hearing loss that begins in the high frequencies and is progressive with age[266]. Therefore, it is a challenge to comprehensively assess the prevalence of ARHI in a population sample, meaning that in this chapter, when assessing the UKBB hearing data for use as a phenotype that is an indicator of ARHI there are several limitations to consider. The SIN test and questionnaires do not assess the audio frequency of a participant's hearing loss. Secondly, the majority of the data is taken at one time point and so the progressive nature of all individuals' hearing abilities cannot be assessed. With the available UKBB data therefore, the phenotype that 'best' describes ARHI will not be based on these factors.

As noted, any bias needs to be limited when defining a trait and its prevalence in a population sample[264,265]. By using the UKBB, some of these potential sources are eliminated already; cases and controls will be selected from the same sample (the whole UKBB sample) and the assessment of hearing impairment is the same for all individuals. In addition, participants were not selected to take part in the study based on hearing-related factors; unselective recruitment was conducted.

Several studies have been conducted using the UKBB hearing-related data fields since the initial data release in 2014. The first study presented a 'snapshot' of the data; the report stated that 1/10 participants experience a 'substantial' hearing impairment based on the results from the SIN test and highlighted the underutilization of hearing aids[123]. Subsequent publications have studied the hearing data in relation to other traits such as vision[267], cognition[122] and depression[195] along with factors such as socio-economic status[268], ethnicity[123] and education[267]. The initial report on hearing-related phenotypes in the UKBB used the hearing test score (SRT) for the best-performing ear to classify participants; 'normal hearing' (SRT < -5.5dB), 'insufficient hearing' (-5.5dB < SRT < -3.5dB) and 'poor hearing' (SRT > -3.5dB). Unexpectedly, this study reported that males had a slightly reduced risk of 'insufficient' hearing impairment and there was no difference in risk for male or females in the 'poor' hearing category. Authors suggest that this either indicates that the increased male susceptibility to hearing loss reported in some studies is due to modifiable risk factors, or that this particular SIN test is not sensitive enough; speech recognition may be unaffected due to the high redundancy of the speech signal[123].

A later study on the relationship between functional hearing (SRT score) and four UKBB cognitive measures reported an unexpectedly high proportion of 40-50-year-olds within the UKBB sample with 'normal' hearing[122]. This study also refers to an internet-based SIN test which has since been developed that has a greater sensitivity to high frequency hearing loss than the UKBB SIN[269,270]. While some of the same UKBB data fields will be assessed, the analyses summarised above will not be directly replicated or expanded on in the work presented in this thesis. Here, the aim is to (1) identify a viable, reliable phenotype that is a suitable indicator of ARHI within the cohort and (2) that can be used as a phenotype for genetic association analysis.

3.2 Results

In the following sections the UKBB SIN test scores (SRT scores) and hearing questionnaire responses are analysed as potential surrogate phenotypes for ARHI genetic association analysis.

3.2.1 SIN test data in the UKBB sample

SRT scores have not previously been used as a phenotype measure for association analysis and the UKBB speech in noise test is the first speech in noise test to be used on a population of this size. Therefore, prior to conducting association analysis, various methods were used to assess the validity of SRT as a surrogate phenotype for ARHI, the reliability of the data and, subsequently, its' suitability for use as a phenotype in association analysis.

The UKBB SIN test procedure

The UKBB SIN test is a form of the 'Digit Triplet Test'. The UKBB version was adapted for use from the original Smits Digit Triplet Test[188] that is introduced in chapter 1, section 1.3.4. Adaptations were made specifically for the UKBB study and are based on work conduced at the University of Southampton[194] and are described in detail by Hall, 2006 which can be accessed via http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=158561. It was then subsequently tested on a normative sample of 20 volunteers with normal hearing, aged 18-29 at the time of assessment (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=138440).

At the UKBB assessment centres, participants were provided with a touch screen monitor and Sennheiser HD-25 headphones with which to complete the SIN test. Participants completed the SIN test twice; once for each ear. Participants with even study ID numbers performed the test on their left ear first and right ear second, and *vice versa* for those with odd study IDs.

The UKBB protocol for the SIN test can be accessed via http://biobank.ctsu.ox.ac.uk/crystal/docs/Hearing.pdf. Briefly, volunteers were asked to set a comfortable volume for each ear prior to beginning the test. They were then played a set of three digits, E.g. '1, 2, 3' and were prompted to enter the digits that they heard, using the touch screen monitor. The test follows an adaptive procedure; the following triplet is played with either an increased or decreased level of background noise, depending on whether the participant correctly or incorrectly identified the digit triplet. After the participant has completed 15 trials of triplets for the first ear, the test for the

second ear begins. The volunteers were asked to set a comfortable volume for the second ear, and 15 trials of triplets were then played.

For each of the 15 trials, the participant will have a 'speech to noise ratio'. This describes the noise level ratio between the digits played and the background noise. The ratio of speech to noise at the first trial was either -2dB or +2dB. Originally, all participants began the test with a speech to noise ratio of -2dB at the first trial. Part-way through the recruitment period, the protocol of the test was changed such that the speech to noise ratio of the first trial is instead +2dB so that so that it was easier for individuals that are hard of hearing to begin the test. It is stated that the protocol was implemented from August 2009.

Upon completion of the test, each participant will have 15 SNR values for each ear (a speech to noise ratio for each of the 15 trials). By convention, the final score for the hearing test is the mean of the SNR values for the final 8 trials and is called the 'Speech reception threshold' (SRT). Each participant thus has one SRT score for each ear. The 'highest' possible score for this test is -12dB (very good hearing) and the 'lowest' possible score is +8dB (very poor hearing). Also by convention, the 'best ear' SRT score per participant is used for analysis. This minimises the possibility of observing effects due to significant environmental exposures (such as past trauma or infection) that have affected the function of just one ear. Thus, for each individual, the 'best ear' SRT score (SRT_B) was the lowest score for either the right SRT, (SRT_R) or left SRT (SRT_L), UKBB data fields as listed in - Methods 2.1.1.1. The SRT_R and SRT_L scores are however useful measures to assess the prevalence of bilateral hearing deterioration in the sample and, or the reliability of the measure within each participant.

SIN test score	Definition	
SNR	Speech to noise ratio	This describes the noise level ratio between the digits played and the background noise Each individual has 30 SNR values. 15 for each ear, for each trial in the SIN test.
SRT_L	Speech reception threshold, left ear	Mean value of the final 8 SNR values for the left ear. Each participant has one SRT_L
SRT_R	Speech reception threshold, right ear	Mean value of the final 8 SNR values for the right ear. Each participant has one SRT_R
SRT_B	'Best' Speech reception threshold	For each individual, the SRT_L or SRT_R is used as the SRT_B value, depending on which had the best score. (-12 best possible score, +8 worst possible score)

Table 3.1. Definitions for the acronyms used to describe types of SIN test scores. The UKBB data field IDs corresponding to these scores are listed in Section 2.1.1.1.

Assessing the normality of the SRT distribution for a quantitative distribution

The SRT score is a quantitative measure. Genetic association analysis of quantitative measures is performed using a linear mixed models approach and so is constrained by the assumption that a quantitative trait is normally distributed. Therefore, to assess whether the raw SRT scores were a suitable phenotype for association analysis, the SRT scores for each participant were plotted in Figure 3.1 and assessed for normality (distribution skewness, Table 3.2). It is not possible to perform the standard Shapiro-Wilk test as it is not suitable for data samples with n>5000 (the greater the number of data samples used in the calculation, the greater the chance of a rejection of the null hypothesis). In the absence of a Shapiro-Wilk test, the normality of the data can be estimated based on the distribution skewness. Skewness that is greater than twice the standard error of the skewness signifies a non-normal distribution.

A distribution with a skew >1 is determined 'highly skewed'; all three of the $SRT_{R/L/B}$ distributions are highly skewed (Table 3.2) and can be visualised in the histograms in Figure 3.1, which display the distributions of SRT_L, SRT_R and SRT_B for the whole sample. Poorer hearing is represented by greater SRT scores i.e. positive SRT scores. All three distributions are non-normal; the skewness is greater than twice the standard error of the skewness (Table 3.2). As raw SRT_B distribution is highly skewed (>1) it is not a suitable quantitative measure for genetic association analysis in the current format. In order to use the SRT scores for association analysis, a strong transformation would be required, or categorisation of scores to create a binary trait of cases and controls.

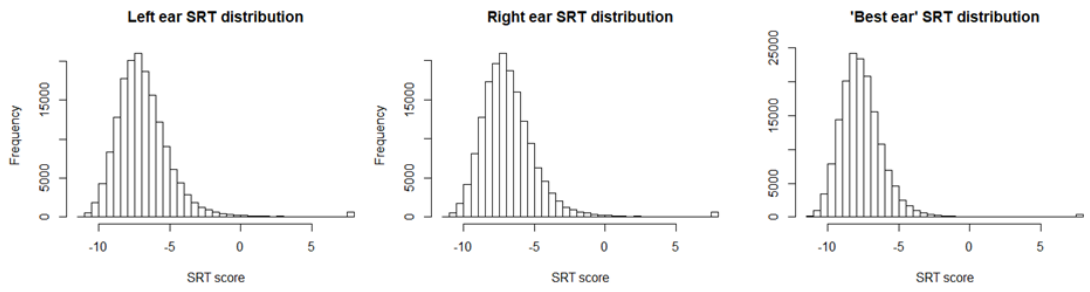

Figure 3.1. SRT distributions for the left ear, right ear and 'best' ear SRT scores in the UKBB sample.

Lower scores signify greater hearing ability; an SRT score of +8 is the 'worst' possible score on the test and signifies a very poor hearing ability. A score of -12 signifies a very good hearing ability. All three distributions are highly skewed to the right; skewness >1 is classified 'highly' skewed. Left ear SRT distribution, SRT_L; Right ear SRT distribution, SRT_R; 'Best ear' SRT distribution, SRT_B.

	N	Mean SRT (sd)	Median SRT	Skewness	SE	IQR	SRT Range	
SRT_L	163296	-6.59 (2.10)	-7	2.39	3.67E-05	2.5	-11.25	+8
SRT_R	163237	-6.55 (2.11)	-7	2.36	3.68E-05	2.5	-11.5	+8
SRT_B	160942	-7.36 (1.70)	-7.5	2.75	3.75E-05	2	-11.25	+8

Table 3.2. Summary of SRT scores in the UKBB sample.

SRT_L; SRT scores for the left ear, SRT_R; SRT scores for the right ear, SRT_B; SRT scores for the 'best' ear, calculated as the lowest SRT score for each individual. Skewness, distribution skewness; SE, standard error of the distribution skewness; IQR, interquartile range of the SRT score distribution. All three distributions are highly skewed to the right.

Defining SRT thresholds to create a qualitative phenotype

Previous studies have used SRT score thresholds to categorise participant hearing ability as either good, insufficient or poor and are calculated such that 'The 'insufficient' category is performance lower than -2 standard deviations with respect to the normative sample while the 'poor' category is defined by a further 2 dB step'[123,188]. This follows a standard statistical classification of 'normal' data points within a sample being within 2 standard deviations of the sample mean[264]. Here the 'normal' proportion is denoted 'good' hearing ability.

Figure 3.2 displays a previous example of this for a different SIN test in a different sample. It displays a distribution of $SRTT_n$ scores (equivalent to the UKBB SRT scores) from the Smits Digit Triplet Test[88] marked with classifications of hearing ability. The Smits test is the test that the UKBB SIN was adapted from. $SRTT_n$ scores from 39,968 participants that completed the Dutch telephone SIN test are displayed. Based on these

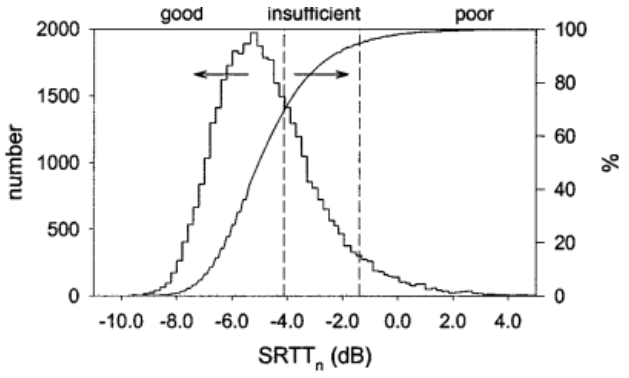

Figure 3.2. SRT distribution from Smits 2005[88] in an elderly Dutch sample.

This figure displays a 'histogram and a cumulative histogram of SRTTns in 0.2-dB intervals. Vertical dotted lines depict borders between the different test results in terms of good, insufficient, and poor.' The SRTTn measure for this Smits Digit Triplet telephone screening test is equivalent to the UKBB SIN test SRT. The sample includes individuals aged >18, with 75% of the sample >44 years of age. Marked on the histogram are the classifications of hearing ability determined by SRT score thresholds; good SRT <4.1 dB, insufficient SRT -4.1dB <= -1.4dB and poor SRT > -1.4dB.

classifications, 67% of the Smits sample have 'good' hearing, 26% have 'insufficient' hearing and 6% have 'poor' hearing.

This method can also be used to categorise the UKBB SRT data. A normative sample consisting of 20 volunteers (aged 18-29 years with normal hearing) was used to derive classification thresholds devised for the UKBB data; 'normal hearing' SRT < -5.5dB, 'insufficient' hearing SRT -5.5dB to -3.5dB and 'poor' hearing SRT > -3.5dB. Figure 3.3 displays the distribution of the SRT_B scores from the UKBB sample, with the UKBB classification thresholds[123,271] overlaid in purple. According to the UKBB classification thresholds, 11.92% of the UKBB SRT_B scores are here 'insufficient' or 'poor'. Based on these classifications, the prevalence of insufficient or poor hearing in the UKBB sample is <12% (Figure 3.3). This is in contrast to the Smits sample, where 32% had poor or insufficient hearing based on the corresponding thresholds (Figure 3.2).

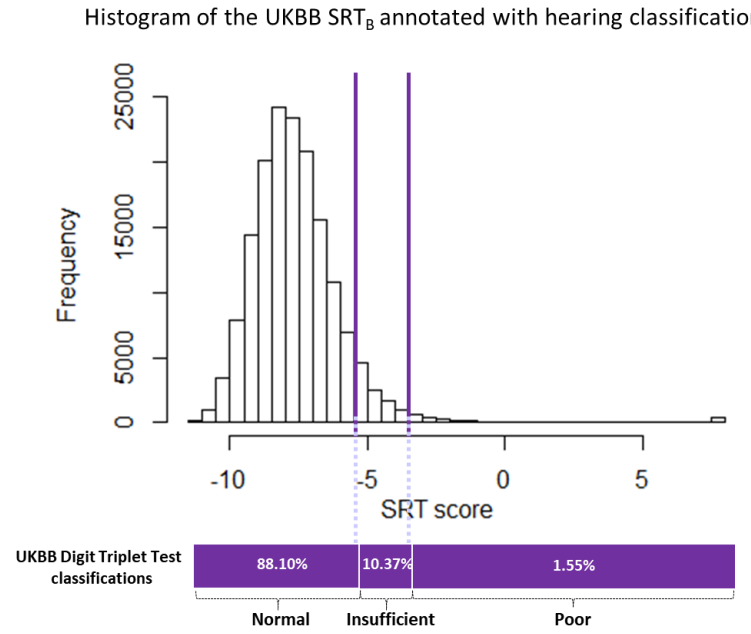

Figure 3.3. Histogram presenting the SRT_B distribution for the UKBB sample.

The histogram distribution is the same as presented for SRT_B in Figure 3.1. Classifications of hearing ability according to the UKBB classification thresholds Dawes 2014[123]are marked below the histogram. The proportions of the UKBB sample that falls into each of these categories are represented by percentage values in the bar shaded purple. Vertical purple lines indicate these proportions on the histogram distribution.

An explanation for the difference in prevalence (12% in UKBB and 32% in the Smits sample) is that different SIN tests vary; as discussed previously the UKBB test has been adapted from the original Smits Digit Triplet Test. For example, the Smits Digit Triplet test was conducted via telephone in a different language to the UKBB test, and different adaptive procedures are used for each of the two tests.

The population samples also vary, making comparisons between the studies difficult. The age distributions in the Smits and the UKBB samples differ; the Smits sample consists of participants aged >40, of which 75% are >44 years old. The UKBB sample participants are aged 40-69. Taking this into account, we may expect the prevalence to be lower in the Smits sample as there are younger participants in this sample, yet this is the opposite of what is observed with these threshold categories. This indicates that the UKBB SIN test is either not as sensitive as the Smits Digit Triplet test, that the prevalence of hearing impairment is genuinely 20% greater in the Smits sample, or that classification method used, incorrectly defines the true prevalence for at least one of the two samples.

3.2.1.1 Is SRT_B a valid indicator for ARHI in the UKBB sample?

As the prevalence of hearing impairment in the UKBB sample is lower than expected, analysis was conducted to assess whether the SRT accurately measures ARHI in the sample (a progressive hearing loss with increasing age, higher prevalence in male subjects and presents as a bilateral HL[266]). The sample was studied to determine whether the prevalence of poor SRT scores increases with age and whether the test scores indicate a bilateral hearing deterioration.

The relationship between SRT_B and age and sex

To assess the relationship between the SRT score and age, a Spearman's rank correlation between SRT_B and age in years was calculated; $r^2_s = 0.238$, $p<2.2e-16$ (Figure 3.4). The correlations between age in years and each ear individually were 0.216 (SRT_L) and 0.217 (SRT_R). This indicates that there is a positive correlation with age; as age in years increases the SRT score increases, yet the correlation is modest; 0.238, Figure 3.4. A study conducted in 2017 on a group of participants aged >50 years observed correlations of age $r^2=0.317$ for men and $r^2=0.354$ for women, where hearing ability was tested with PTA using the WHO classification thresholds[96].

As these previous correlations were based on PTA data, they are not directly comparable to the UKBB SIN data. No previous study has reported an SRT score correlation with age. Therefore, the SRT scores were plotted by age group and sex, alongside data from the Smits Digit Triplet test (Figure 3.5).

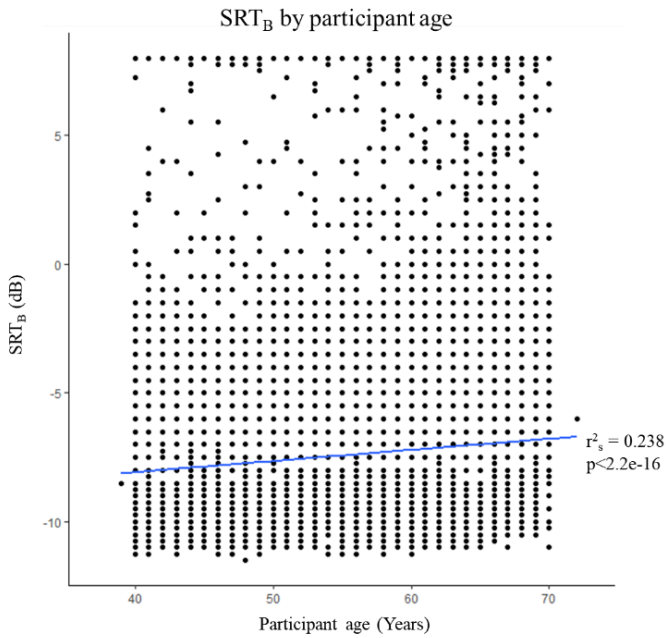

Figure 3.4. Scatter plot of SRT_B scores (dB) by participant age (years).

The correlation of SRT_B and age is marked on the graph; r2s = 0.238, p<2.2e-16. The regression line is marked in blue.

Figure 3.5 is presented as a line graph that displays the mean SRT scores between different age groups for the two study samples, split by sex. The red and green lines on Figure 3.5 are plotted with the mean SRT_B values for the female and male samples in the UKBB sample. The blue and purple lines are plotted with the mean $SRTT_n$ values from a Smits Digit Triplet Test sample[272]. Table 3.3 displays the raw data for Figure 3.5. Both Figure 3.5 and Table 3.3 demonstrate that the variability between SRT scores and age and sex is greater in the Smits Digit Triplet Test[273] than the UKBB SIN Test. In the UKBB sample, there is no significant difference between male and female SRT_B distributions (p-value = 0.065).

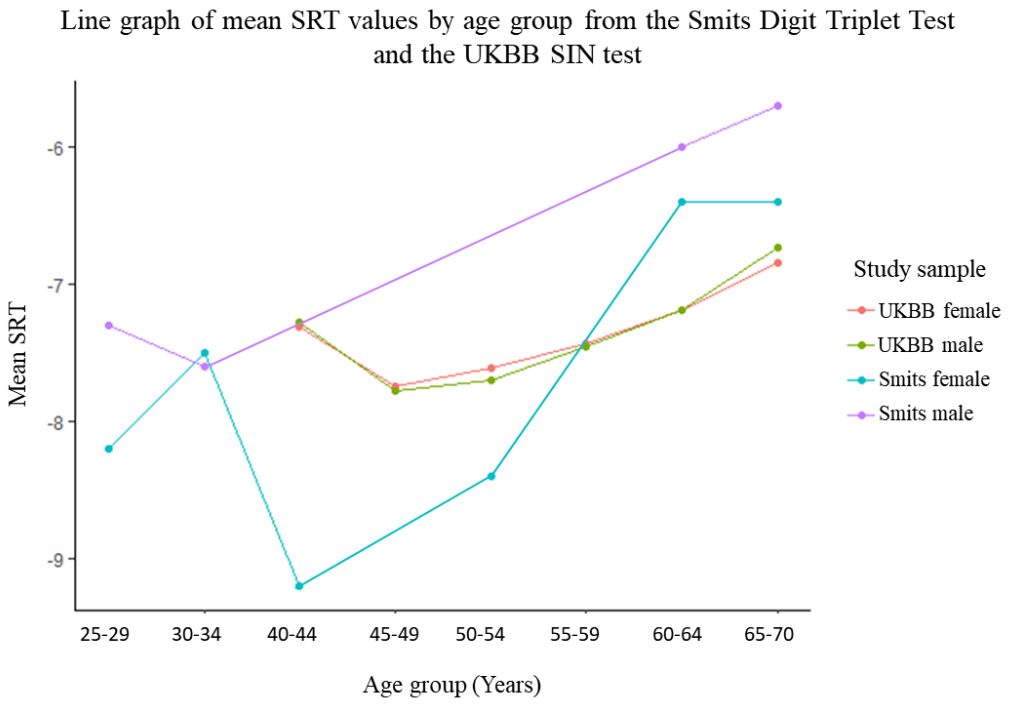

Figure 3.5. Line graph displaying the mean SRT value by age group split by sex for the UKBB sample and Smits Digit Triplet Test sample[273].

The UKBB recruited volunteers aged >39 and so there is no data for the age groups 25-29 and 30-34 for the UKBB sample. There was no data available for some of the age groups in the Smits sample; males aged 40-59 or females aged 30-34 and females aged 55-59.

	UKBB SIN Test				**Smits Digit Triplet Test**			
	Female		Male		Female		Male	
Age group (years)	Median	Mean	Median	Mean	Median	Mean	Median	Mean
25-29	NA	NA	NA	NA	-8.3	-8.2	-7.6	-7.3
30-34	NA	NA	NA	NA	-7.6	-7.5	-8.0	-7.6
40-44	-8	-7.31	-8	-7.28	-9.2	-9.2	NA	NA
45-49	-8	-7.74	-8	-7.78	NA	NA	NA	NA
50-54	-8	-7.61	-8	-7.70	-8.4	-8.4	NA	NA
55-59	-7.5	-7.43	-7.5	-7.46	NA	NA	NA	NA
60-64	-7.5	-7.19	-7.5	-7.19	-7.2	-6.4	-6.2	-6.0
65-70	-7	-6.85	-7	-6.74	-6.8	-6.4	-6.2	-5.7

Table 3.3. Median and mean SRT scores for age groups, split by sex for both the UKBB SIN test sample and the Smits Digit Triplet Test.

These scores correspond to the pot in **Figure 3.5**. The two studies did not use all of the same age groups for analysis. 'NA' is noted where there is no data available for an age group in a particular sample.

Do the SRT scores report bilateral hearing impairment?

ARHI is understood to present as a bilateral hearing deterioration[266]. If the SRT scores accurately represent ARHI, an individual's SRT_L and SRT_R would therefore be highly correlated. Across the sample, the spearman's rank correlation coefficient of SRT_L and

SRT_R is r^2_s=0.358, p-value< 2.2e-16, visualised in Figure 3.6a. Few studies report on the symmetry of hearing loss in a study sample. In a recent study however, 14.8% of males and 13.3% females exhibited asymmetry defined as an interaural difference of more than 10 dB in hearing levels averaged over 0.5, 1, 2, and 4 kHz[96]. As a different type of hearing test was used to assess participant hearing ability, the asymmetry cannot be directly compared between this sample and the UKBB sample. However, the work[96] could support a theory that the unilateral hearing loss observed here in the UKBB SRT scores may reflect unilateral hearing deterioration rather than the low test re-test performance.

Within the whole sample, the median difference between SRT_L and SRT_R was 1dB, and the mean difference was 1.55dB, yet the range was 0-18dB. Figure 3.6b is a second scatter plot of the same data (SRT_L vs SRT_R), but where the colouring denotes the density of points at specific score combinations. The plot enables a greater interpretation of the average differences between SRT_L and SRT_R, the range and the correlation. The majority of individuals cluster around the mean SRT scores ~7dB for both ears (Figure 3.6b) which may be a marker that the test lacks sensitivity. The range of difference between SRT_L and SRT_R for some participants (marked on Figure 3.6 (b), by blue shading and black points) may indicate that either the SIN test was not performed accurately for these participants or that there is a greater prevalence of unilateral hearing deterioration in ARHI sufferers than is currently understood.

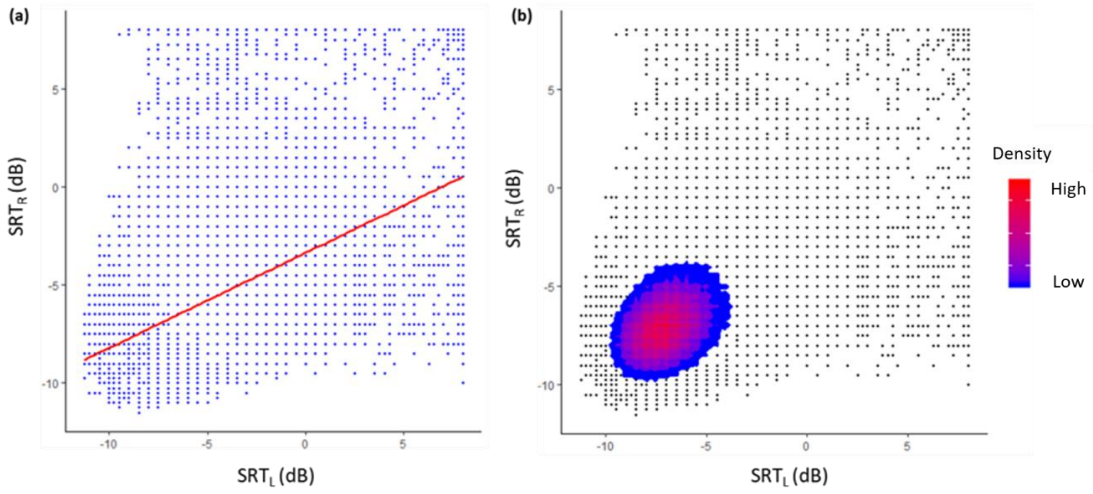

Figure 3.6. Scatter plots of SRTL and SRTR scores for all participants that have an SRT score for both ears.

(a) is a scatter plot of SRT_L by SRT_R r2s=0.358, p< 2.2e-16, with a regression line marked in red. (b) is a density scatter plot of SRT_L and SRT_R. The density of participants with specific score combinations is denoted by colours red to blue and black. This scatter pot demonstrates both the cluster of individuals that have SRT_L and SRT_R scores clustered around the mean values and also the range difference between SRT_L and SRT_R; 0-18dB.

Is a learning effect observed when performing the test on the second ear?

89% of the sample have a different SRT_R score to their SRT_L. While this could be due to genuine differences in hearing abilities between participants' ears or the measurement error on the test, it may be caused by a learning effect from the SIN test. If a learning effect was observed, it would be more likely that the second ear to be tested would be the 'best' SRT score. This would indicate that an individual is more likely to achieve a 'better' score on the test having completed and 'learnt' how to perform the test, resulting in a disparity between the SRT_L and SRT_R scores.

In the UKBB SIN test, the order of testing for left and right ears was determined by whether the participant study ID was an odd or even number. ~50% of participants performed the test on their left ear first, followed by the right, and *vice versa*. Those with an even study ID number tested the left ear first and those with an odd study ID number tested the right ear first. Table 3.4 displays the number of participants that had the 'best' ear score for left/right ears and which ear was tested first. The proportion of left and right SRT scores that were the 'best' SRT score is not dependent upon which ear was tested. Interestingly, for both groups there is a slight preference for $SRT_L = SRT_B$.

	$SRT_L = SRT_B$ (n)	**%**	**$SRT_R = SRT_B$ (n)**	**%**	**$SRT_L = SRT_R$ (n)**	**%**
Left tested first (even study ID)	36492	45.56	35253	43.39	8946	11.05
Right tested first (odd ID)	36559	45.22	34821	43.69	8871	11.09

Table 3.4. Comparison of SRT_L and SRT_R scores in relation to the order of testing.

The table displays the number of participants that that achieved a 'better' (lower) SRT score for their left ear ($SRT_L = SRT_B$), a 'better' score for their right ear ($SRT_R = SRT_B$) or the same SRT score for both ears ($SRT_L = SRT_R$). The proportion of each of these combinations are reported as percentage of the sample that either had the right or left ear tested first.

3.2.1.2 Assessing the reproducibility of the SRT score

Results in the previous section indicate a higher prevalence of unilateral HL than would be expected. While this may be the case, it contests the understanding that ARHI presents as a bilateral deterioration. Here I present work to explore whether these results are due to the insensitivity of the test and the reliability of the measure. The test re-test reproducibility of a measure indicates its validity and stability. The UKBB SIN was not

included in a pilot study and so the reproducibility of the measure has not previously been determined.

Due to the way that the UKBB data were collected, there are two possible ways to study the reliability of the SRT test. Firstly, 10 UKBB assessment centres conducted the SIN test during the initial phase of recruitment (Birmingham, Bristol, Croydon, Middlesbrough, Nottingham, Sheffield, Liverpool, Hounslow, Swansea and Wrexham). To determine whether the test was reproducible in these different environments, i.e. the SRT distributions were consistent across different assessment centres, samples from each assessment centre are analysed based on (1) SRT distribution and (2) SRT correlation with age.

Furthermore, a subset of individuals (n=4544) attended the Cheadle assessment centre for a repeat visit 2-4 years after their initial visit, and there completed a SIN test. With this sample, it is possible to test the reliability of the SRT score over time. A reliable measure would have a high correlation of scores between the first and second assessment over this short time period.

Are the SRT distributions uniform across the different UKBB assessment centres?

The SIN test was conducted at 10 of the UKBB assessment centres during the main phase of recruitment. Table 3.5 contains summary figures of the SRT_B scores (and age of participants) at each of the 10 centres and Figure 3.7 displays the individual SRT_B distributions. A Kruskal-Wallis rank sum test confirms that the distributions of best SRT scores are significantly different between centres, Kruskal-Wallis chi-squared = 7254.3, df = 9, p-value < 2.2e-16, (critical value of 16.9190 for 9df and alpha 0.05), indicating that it may not be a reproducible measure. However, it is challenging to discount external factors impacting SRT scores at each centre or inherent differences between the sub cohorts such as age distributions.

Assessment centre	N	Mean SRT_B	Median SRT_B	Mean Age	Median age
Birmingham	23370	-7.23	-7.5	56.29	57
Bristol	11241	-7.14	-7.5	55.54	57
Croydon	25956	-7.03	-7	56.51	58
Hounslow	26379	-7.45	-7.5	56.15	57
Liverpool	19418	-7.51	-7.5	57.68	59
Middlesbrough	17936	-7.67	-8	56.67	58
Nottingham	4602	-6.88	-7	57.22	59
Sheffield	29240	-7.70	-8	57.18	59
Swansea	2185	-6.04	-6	57.45	59
Wrexham	615	-6.06	-6.5	56.89	58

Table 3.5. Summary table of the number of participants that completed the SIN test at each UKBB assessment centre during the main phase of recruitment.

The table also displays the mean and median SRT_B scores at each UKBB assessment centre, and the mean and median ages of the participants that completed the SIN test at each centre. A Kruskal-Wallis rank sum test confirmed that the distributions of best SRT scores are significantly different between centres, Kruskal-Wallis chi-squared = 7254.3, df = 9, p-value < 2.2e-16, (critical value of 16.9190 for 9df and alpha 0.05).

To explore the individual SRT_B distributions with respect to age in each sample, box plots are presented in Figure 3.8. These plots display SRT_B scores subset by assessment centre and participant age groups. As expected, all centres display a trend of increasing SRT_B score with increased age. Nottingham, Swansea and Wrexham assessment centres visibly have a higher median SRT score for all age groups than the remaining centres. However, from Table 3.5, it is clear that the sample size in these centres are relatively low (<4,603) compared to all other centres (>17,936). The outliers that are present in these three centres may therefore likely have a greater effect on the median SRT_B for these centres (median denoted by the mid-line of the box plot) and therefore could explain the observed shift.

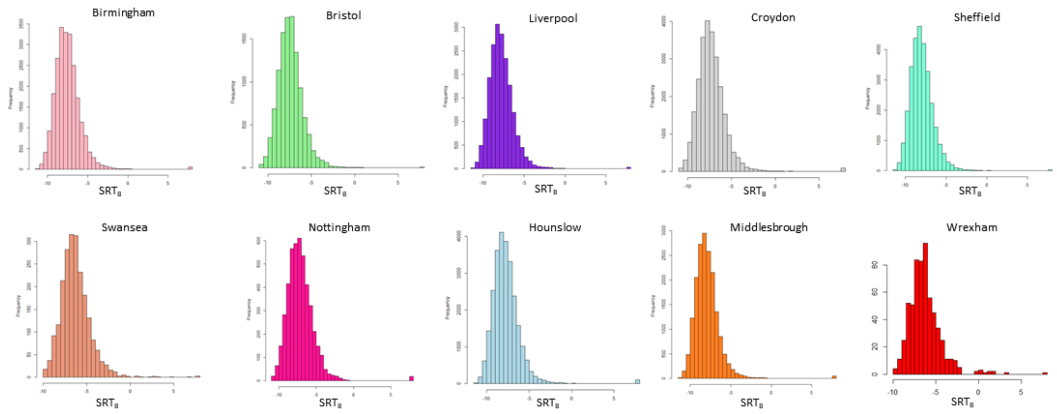

Figure 3.7. Histograms of SRT_B distributions plotted individually for the ten UKBB assessment centres.

Distributions are for samples from the ten centres that ran the SIN test during the main phase of recruitment. The frequency (y-axis) scale differs between the plots as the sample size for each centre differs (Table 3.5). X-axis is labelled SRT_B.

In addition, Figure 3.8 demonstrates that while there is a trend for the SRT_B score to increase as age increases (corresponding to the expected hearing deterioration with increasing age) within each centre, the median score may not be comparable between centres. For example, while the median SRT_B score increases with age in both the Wrexham and Sheffield samples, the lower interquartile range of the Sheffield SRT_B 65-70 age group is the same as the SRT_B median in the Wrexham <45 age group. This is another indicator of irreproducibility at different sites, or intrinsic differences between samples which impact the SRT_B scores. This could be due to the effects of a number of hypothesised ARHI-related risk factors. Without determining these factors and the effects on the SRT score, they cannot be accounted for and it may not be suitable to analyse the data as one sample.

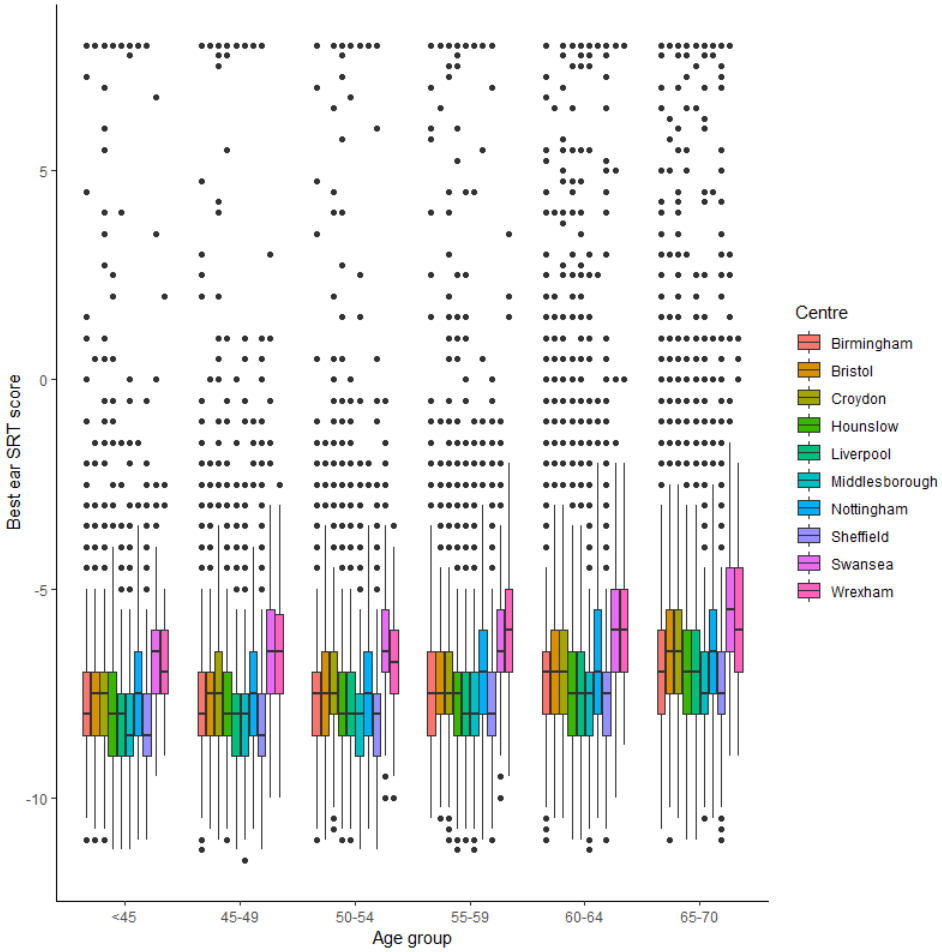

Figure 3.8. SRT_B scores displayed as boxplots according to UKBB assessment centre and age group.

The boxplots display the median SRT_B for each age group (split by assessment centre) and the interquartile range (25th percentile, Q1 and 75th percentile, Q3). The maximum (Q1-1.5*IQR) and minimum values (Q3+q.5*IQR) are denoted by the whiskers and the outliers are presented as black dots outside of these minimum and maximum values.

Did the protocol alteration affect the SRT distributions?

A protocol alteration during recruitment consisted of changing the speech to noise ratio (SNR) at the first trial to +2dB, instead of -2dB. This was so that participants who were hard of hearing would more easily be able to begin the test procedure (see link to the UKBB protocol, p.85). The SIN test was carried out over different time periods in each assessment centre. Therefore, the protocol change in August 2009 will have had a different impact on different assessment centres, depending on when the hearing test was running at each centre. This could therefore be an external source of the variation that is observed in SRT scores between assessment centres demonstrated in Figure 3.8. In total, 49,686 participants completed the SIN test with the original protocol and 116,642 completed the SIN test with the adapted protocol. Figure 3.9 displays the dates

of SIN testing at each assessment centre and the colour denotes which protocol was in use.

An analysis of the raw UKBB SIN data shown in Figure 3.9 did not meet the expectation that the SIN test protocol change occurred in August 2009, as stated in the UKBB protocols (see link to UKBB protocol p.85). This discrepancy was discovered by plotting a visualisation of the SIN ratio at the first trial for each participant (either +2dB or -2dB) and subsetting the data by assessment centre and by date of assessment, Figure 3.9. It is apparent that not all the assessment centres adapted the new protocol on the intended date of August 2009. The Nottingham, Middlesbrough and Bristol assessment centres did not adapt the new protocol at all, while Sheffield, Liverpool and Hounslow adopted the protocol at different time points. Four assessment centres began SIN testing after August 2009. Of these, Wrexham, Swansea and Birmingham all used the new protocol only, while 230 participants at Croydon performed the SIN test with the old protocol before the new version was implemented. Due to the very small number of participants (230), this is difficult to visualise in Figure 3.9.

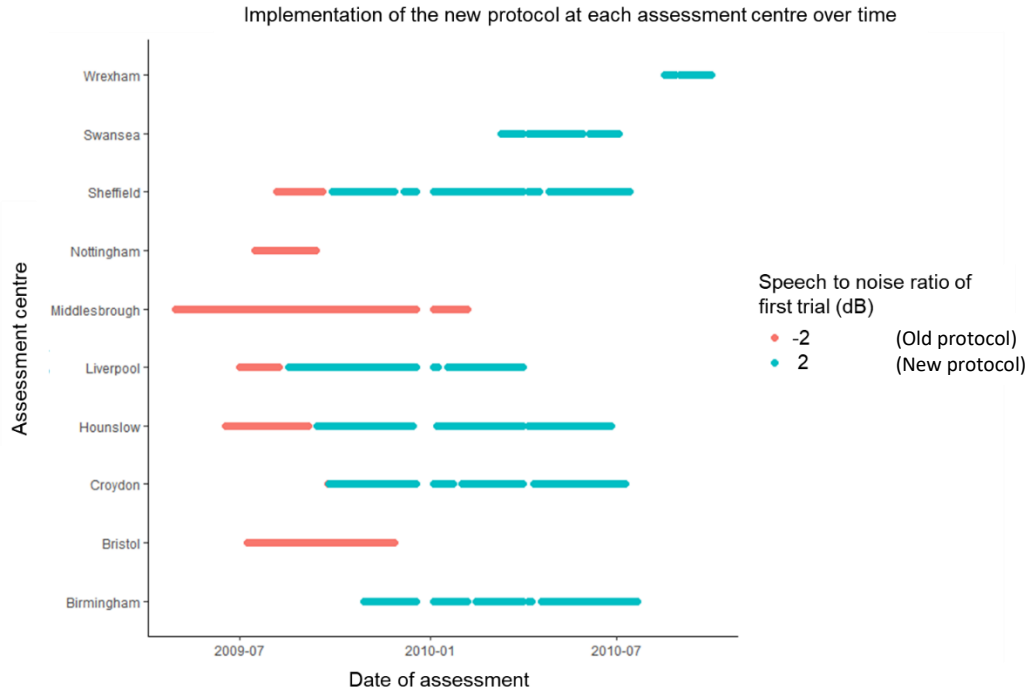

Figure 3.9. Graph displaying the dates that the SIN test was conducted at each assessment centre and which protocol was being used.

Orange represents the testing period with the original protocol (first trial SIN -2dB) and turquoise represents the new protocol (first trial SIN +2dB). From this graph, it is apparent that not all of the assessment centres adopted the protocol change at the same time, August 2009 as the UKBB protocol states.

Table 5 displays the number of participants at each of the assessment centres and the exact dates that testing commenced and was completed. A Man-Whitney-Wilcoxon test

indicates that the SRT distributions are significantly different under the two protocols; W = 2.751+09, p-value < 2.2E-16. However, this significant difference could be due to the underlying significant difference between the assessment centres (identified in section 1.1.3), rather than solely the protocol change.

To explore whether this is due to the protocol change rather than the differences between assessment centres, the calculation was performed individually on the centres that used both protocols: Sheffield, Liverpool, Hounslow and Croydon. This removes the effect of the factors that cause variability between assessment centres. The SRT_B distributions of the two protocols were significantly different in Sheffield, Liverpool and Croydon ($p<0.05$) but not in Hounslow.

Assessment centre	N for each protocol		Mann Whitney test for SRT_B between protocols
	-2	+2	
Bristol	11604	0	-
Nottingham	4719	0	-
Sheffield	3815	25928	W = 50334000, p-value = 8.107E-11
Liverpool	3317	16439	W = 24670000, p-value = 7.268E-06
Middlesbrough	19796	0	-
Hounslow	6206	21054	W = 58770000, p-value = 0.01622
Croydon	230	26243	W = 2307600, p-value = 1.939E-06
Birmingham	0	24138	-
Swansea	0	2212	-
Wrexham	0	630	-

Table 3.6. A comparison of SRT_B scores within assessment centres under different SIN protocols.

Table listing the number of participants that completed the SIN test under each protocol, at each assessment centre, and the results of Mann Whitney test for difference between the SRT_B distributions under different protocols, within the four centres that ran both of the protocols.

Are participant SRT scores consistent over time?

Having assessed the reproducibility of the test between different centres, the reproducibility was also tested over a period of time. 4529 individuals completed the SNR test twice; once during the initial recruitment phase (2006-2010) and again during the first repeat assessment (2012-2013). This subset can be used to calculate the test-re-test correlation. The period of time between the first (T0) and second (T1) assessments differed between participants; the range in time between first and second visits was 2.11-3.88 years (770-1417 days). The mean and median number of days between tests were 1103.98 and 1100 respectively.

For a reliable test, SRT_B difference between the two assessments would be negligible for the vast majority of participants over this short time period. The correlation of SRT_B scores between the first and second visits was r^2=0.27, p<2.2E-16, graphically displayed in Figure 3.10 (SRT_L r^2=0.25, p<2.2E-16 and SRT_R r^2= 0.23 p<2.2E-16). There was a significant difference between the T0 and T1 scores; p=7.26E-11 with a mean difference of 0.17dB (sd.=1.65) between T0 and T1, unadjusted for the time elapsed between T0 and T1. The distribution of SRT_B scores at T0 and T1 is displayed in the density plot in Figure 3.11, a shift of SRT scores to the right on the distribution is visible for T1.

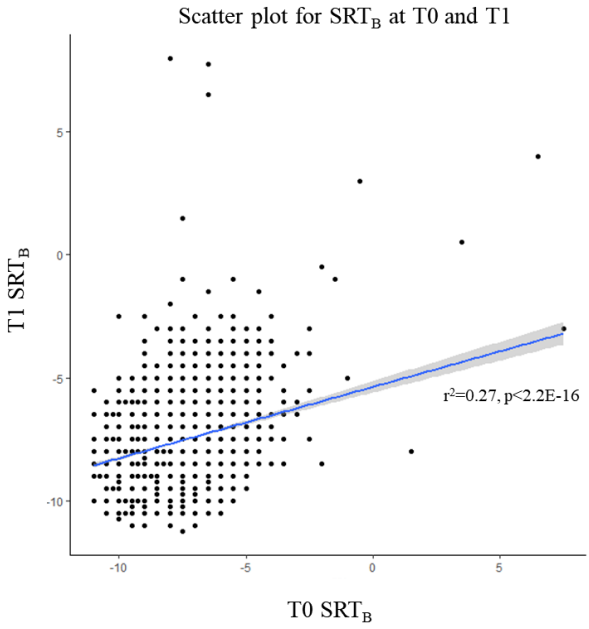

Figure 3.10. Scatter plot for SRT_B scores at T0 and T1.

The correlation is 0.27, the regression line is marked in blue with the standard error shaded in grey. T0, SRT_B score at baseline; T1, SRT_B score at the repeat visit. The black dots on the graph do not comprehensively depict the density of data points. One black dot can represent multiple datapoints at each score combination. This means that without the regression line present, the points alone indicate a correlation closer to 1.

This relatively low correction and significant difference between T0 and T1 scores could be explained by the natural deterioration of participants' hearing over time. If the natural deterioration of hearing is the cause of the low correlation, there would be a correlation observed between the SRT change over time and the length of time between assessments; the greater amount of time that has elapsed, the greater the expected change in SRT. The mean difference in SRT_B score per year elapsed was 0.05dB, (sd=0.56) and the Pearson's product-moment correlation between time elapsed between assessments and SRT_B is 0.06 p<2.2E-16. Both estimates indicate that there is a slight deterioration in hearing over time within the sample. The low correlation could also

signify the alternative sources of variation or poor test-re-test reliability, as discussed further in the discussion section of this chapter.

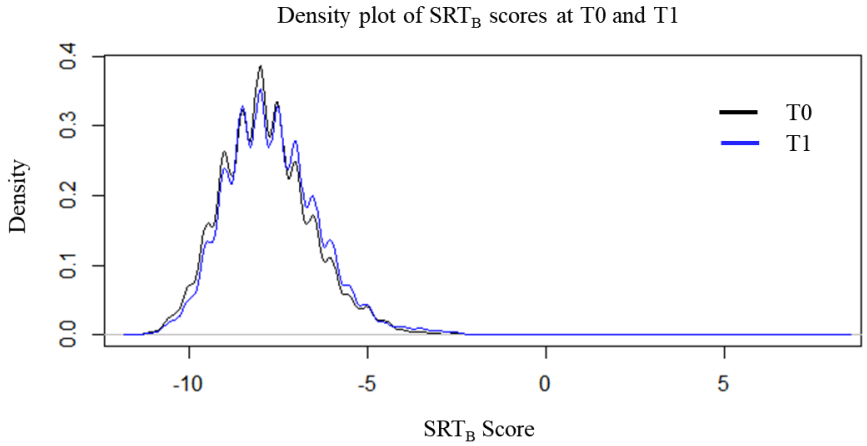

Figure 3.11. Density plot of SRT_B scores for T0 and T1 for the subset of individuals that completed the test twice. n=4529

The black line marks the density distribution of SRT_B scores at T0 and the blue line marks the density distribution for SRT_B scores at T1. Here the shift (mean difference of 0.17dB) between T0 and T1 scores is visible; a shift to the right indicates a shift toward greater SRT scores and thus poorer hearing.

	UKBB	**Smits**
Age group at T0	Predicted mean SRT difference over 5 years in dB (sd)	Mean SRTTn difference over 5 years in dB (sd)
40-49	0.25 (2.73)	0.23 (1.98)
50-59	0.32 (2.68)	0.27 (2.10)
60-70	0.22 (2.90)	0.39 (2.68)

Table 3.7. Predicted SRT score differences over a five-year period.

Mean predicted change in SRT_B score over five years calculated for the UKBB sample (n=4529), and the mean change in $SRTT_n$ over five years with a sample that took the Dutch Triplet Digit test, n=427. The prediction of change in SRT over 5 years was calculated from the SRT change relative to the time elapsed between T0 and T1 per participant. The mean changes are displayed by participant age range at T0, the time point of the first measurement.

Figure 3.12 graphically displays this difference in SRT_B between the first and second assessments, and the amount of time elapsed between assessments. Figure 3.12 displays a scatter plot of the difference between SRT_B at T0 and T1 in dB (y-axis), and the time elapsed in days between T1 and T0 (x-axis). The histograms represent the frequency distributions for the difference in SRT_B over time, and the time that between T0 and T1. This scatter plot does not display a strong relationship between the SRT_B difference between visits and the time elapsed between visits (r^2=0.06, p<4.94E-05), though the relative SRT_B change over time is in the range of the change in time seen in the Dutch Triplet Test (Table 3.7). The scatter and histogram on the y-axis visually display the

difference between the first and second SRT_B scores (which in turn displays the variability within individuals on the performance on the test).

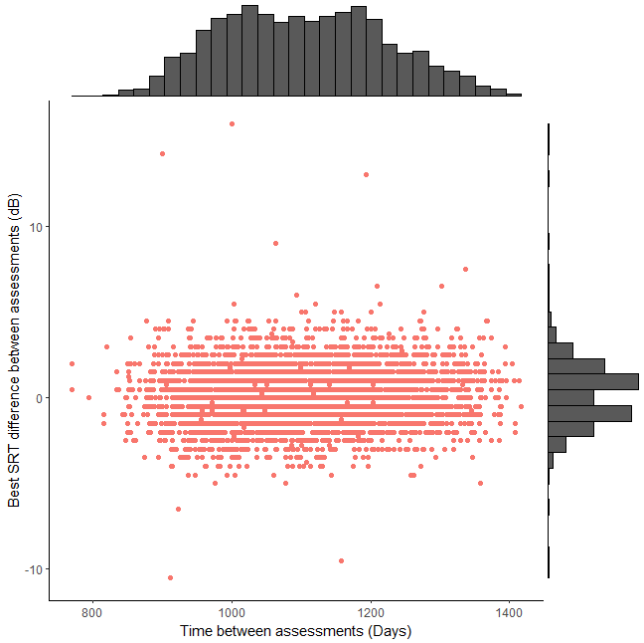

Figure 3.12. Scatter plot of SRT_B difference (dB) between first and second assessments and time between the two assessments (days).

The marginal plots (histograms, grey) display the distribution of time between the two assessments in days (top) and the difference between SRT_B scores for the first and second assessments. The scatter plot visualises both the difference between SRT_B scores between the first and second assessments (right) and the length of time between assessments (top). The scatter plot does not display a strong correlation between time elapsed and increase in SRT_B, r^2=0.06, p<4.94E-05.

3.2.2 Hearing questionnaire measures in the UKBB sample

In addition to the SIN test, participant hearing ability was assessed via questionnaires. The participants completed the questionnaires on touch screen monitors during the assessment centre visit. The three hearing-related questions and the possible responses are listed below in Table 3.8, the UKBB Field IDs are listed in - Methods 2.1.1.1.

Question	Possible responses						
"Do you have any difficulty with your hearing?"	Yes	No	Prefer not to answer	Do not know	I am completely Deaf	-	-
-"Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?"	Yes	No	Prefer not to answer	Do not know	-	-	-
"Do you use a hearing aid most of the time?"	Yes	No	Prefer not to answer	-	-	-	-

Table 3.8 Three hearing-related questions and possible responses that were included in the UKBB questionnaire.

All three of these questions were included in the questionnaires during the initial recruitment period 2007-2010 and in the pilot study. If participants responded, "I am completely deaf" to "Do you have any difficulty with your hearing?" they were not asked either of the two subsequent questions and they did not take part in the SIN test. Until 2009, the question "Do you use a hearing aid most of the time?" was only presented to participants that had answered "Yes" to one of the two questions regarding hearing difficulty. In 2009, when the hearing test was integrated into the assessment day, all participants were asked "Do you use a hearing aid most of the time?" except from those that had responded "I am completely deaf" to "Do you have difficulty with your hearing?" Therefore, due to the restrictions on the question from 2007-2009, only ~3/5 of the sample were asked "Do you use a hearing aid most of the time".

3.2.2.1 Is the hearing aid use questionnaire a suitable indicator for ARHI?

N=307,724 participants responded to the question "Do you use a hearing aid most of the time", UKBB Field IDs are listed in - Methods 2.1.1.1. 4.9% of the sample responded "Yes" to "Do you use a hearing aid most of the time?" while 95% responded "No" (Table 9). These results can be visualised in Figure 3.13, which depicts the differing age distributions between those that responded "Yes" and "No". For both males and females, the age distribution is significantly different between those that responded "Yes" and "No" ($t = 92.023$, $df = 17470$, p-value < 2.2e-16). This indicates that the prevalence of frequent hearing aid use increases with increasing age. There are however a number of outliers (marked by dots below the whiskers). These are present in the younger age groups reporting "Yes", and so may be individuals that have congenital or an early-onset, acquired, form of hearing loss.

"Do you use a hearing aid most of the time?"				
Response	Yes	No	Do not know	Total
N	14966	292192	566	307724
Prevalence %	4.9	95	<1	-
Mean Age	61.87	56.87	56.27	57.11
Median Age	63	58	58	59
Male %	55	48.39	48.59	48.7

Table 3.9 Summary of participant responses to the question regarding hearing aid use.

N, the number of participants with a specific response; Prevalence, the percentage of participants for each response from the total that were presented with the question; Male %, percentage of participants that gave that specific response and are male.

This question can be used to classify participants by hearing impairment; those that use a hearing aid most of the time will have been diagnosed with a hearing impairment and subsequently prescribed treatment in the form of a hearing aid device.

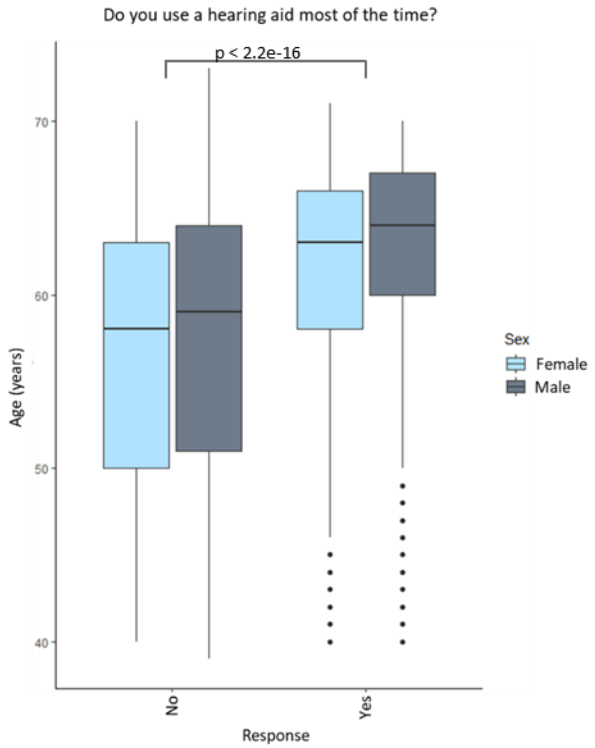

Figure 3.13. Box plots displaying the age distributions of participants that responded "Yes" and "No" to question regarding frequent hearing use.

The boxplots display the median age for each response (split by sex) and the interquartile range (25th percentile, Q1 and 75th percentile, Q3). The maximum (Q1-1.5*IQR) and minimum values (Q3+q.5*IQR) are denoted by the whiskers and the outliers are presented as black dots outside of these minimum and maximum values. The mean age of participants that responded 'Yes' and 'No' was significantly different, the t-test p value are annotated on the plot.

3.2.2.2 Are the hearing difficulty questionnaire measures suitable indicators of ARHI?

The UKBB recruited participants aged >40, and so is a suitable population to use to study ageing traits. However, from these questionnaire measures, the age of hearing difficulty onset cannot be determined. Table 3.10 and Table 3.11 display the counts and percentage of each of the possible responses for the two questions. A quarter of the sample (24.5%) responded that 'Yes' they have difficulty with their hearing, while over a third (36.74%) responded that 'Yes' they have difficulty with their hearing in the presence of background noise. From Table 3.10 and Table 3.11 it is also possible to see that the mean and median ages are higher for 'Yes' respondents than 'No' respondents for both groups.

Do you have any difficulty with your hearing?					
Response	Yes	No	Do not know	Prefer not to answer	I am completely deaf
N	122075	355466	19772	516	130
Prevalence %	24.5	71.39	3.97	0.001	0.00026
Mean age	58.7	55.8	56.48	55.87	56.72
Median age	60	57	58	57	58

Table 3.10. Summary of participant responses to the first question regarding hearing difficulty.

N, number of participants that selected each response; Prevalence %, the percentage of the sample that selected each response.

Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?				
Response	Yes	No	Do not know	Prefer not to answer
N	184290	306683	9957	675
Prevalence %	36.74	61.14	0.02	0.001
Mean age	57.8	55.76	56.42	55.69
Median age	59	57	58	56

Table 3.11. Summary of participant responses to the first question regarding hearing difficulty when background noise is present.

N, number of participants that selected each response; Prevalence %, the percentage of the sample that selected each response.

3.2.2.3 Is there a positive relationship between age and self-reported hearing difficulty?

The prevalence of hearing difficulty within the different age groups was investigated further, to establish whether these measures are likely capturing progressive, adult hearing loss such as ARHI, or congenital and early-onset hearing impairments. The age distributions for the individuals that responded "Yes" and "No" to the two questions regarding hearing difficulty can be visualised in the boxplots in Figure 3.14. The age distributions are significantly different (p-value < 2.2e-16) those that responded "Yes" and "No" to the question "Do you have any difficulty with your hearing?" There was also a significant difference between the age distributions of individuals that responded "Yes" and "No" to "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" ($t = 87.785$, $df = 401400$, p-value < 2.2e-16). Participants that responded to "Prefer not to answer" and "Do not know" were excluded from these calculations.

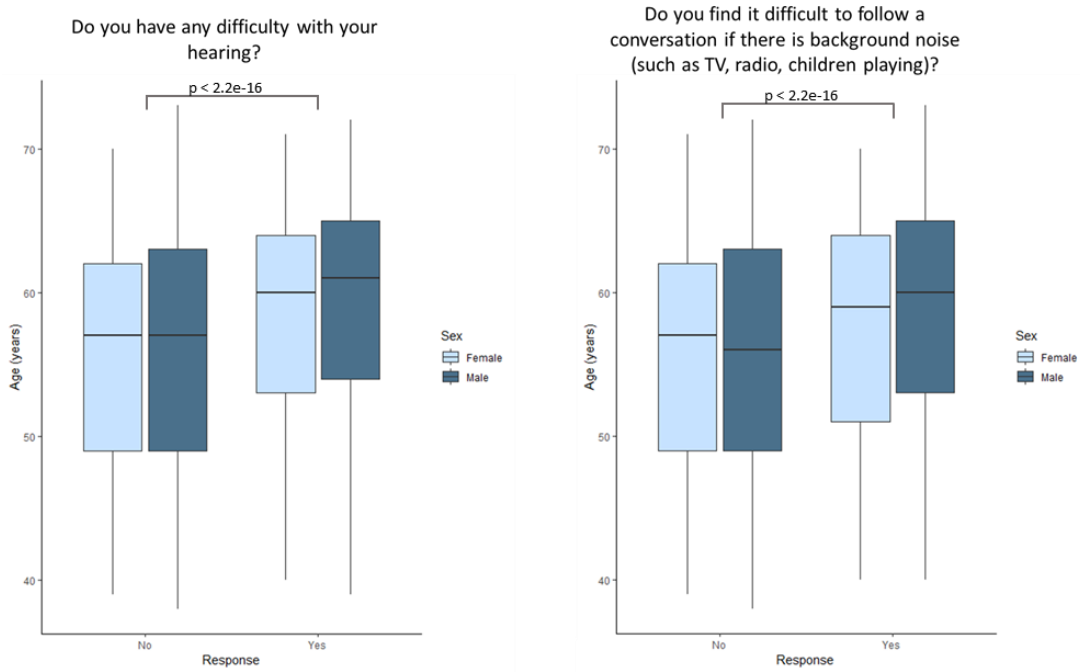

Figure 3.14. Box plots displaying the age distributions of participants that responded "Yes" and "No" to the two questions regarding hearing difficulty.

The boxplots display the median age for each response (split by sex) and the interquartile range (25th percentile, Q1 and 75th percentile, Q3). The maximum (Q1-1.5*IQR) and minimum values (Q3+q.5*IQR) are denoted by the whiskers. The mean age of participants that responded 'Yes' and 'No' was significantly different for both questions, t-test p values are annotated on the plot.

Of the 130 individuals that responded, "I am completely deaf", 65 were male and 65 were female. The mean age for the 130 individuals was 56.72, while the median age was 58, and the age distribution in this sample was not significantly different to the age distribution of the overall cohort, (t = 0.25618, df = 129.07, p-value = 0.7982). This indicates that the portion of the sample that reports being "completely deaf" is more likely to be made up of individuals with congenital or early onset forms of HL, as opposed to ARHI cases. Therefore, this response is likely not suitable to use as a classification of ARHI.

Table 3.12 summarises the combination of "Yes" and "No" responses for the two questions regarding hearing difficulty (HD) and difficulty when background noise is present (HDBN). Responding "No" to both of the questions was the most common combination of responses; 59.62% of the sample responded "No" to both. The prevalence of responding "Yes" to both questions is 21.62%. Only 4.03% of participants experience hearing difficulty without any difficulty when background noise is present, while 15.1% report having difficulty when background noise is present but no general hearing difficulty. This indicates that experiencing hearing difficulty when background noise is present, is more prevalent than general hearing difficulty and may, in some cases,

precede general hearing difficulty. This is of interest for this work as the most common symptom of ARHI is experiencing a difficulty hearing, in particular, deciphering speech when background noise is present. The prevalence of responding "Yes" to both questions is 21.62%, a prevalence that is expected for ARHI in a sample of this age range. In addition, 55.68% of the "Yes" and "Yes" respondents were male, representing a higher male prevalence, as expected in ARHI cases.

Question	**HD**	**HDBN**	**HD**	**HDBN**	**HD**	**HDBN**	**HD**	**HDBN**	**Total**
Response combination	Yes	Yes	No	No	Yes	No	No	Yes	-
N	101642		278509		18953		70970		470074
Age mean	58.94		55.64		57.47		56.40		56.53
Prevalence %	21.62		59.25		4.03		15.1%		-
Age median	61		56		59		58		58
% male	55.68		40.79		51.84		50.26		45.62

Table 3.12. This table summarises the combinations of "Yes" and "No" responses from the two questions regarding participant hearing difficulty and difficulty in background noise.

HD, "Do you have any difficulty with your hearing?"; HDBN, "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" The total number of participants that responded either "Yes" or "No" to both questions was 470,074. All other participants answered either "I am completely deaf", "Prefer not to answer" or "Do not know" to at least one of the two questions. N, number of participants with specific response combination; prevalence %, percentage of participants with specific response combination from the total number of participants that gave one of these 4 response combinations.

3.2.2.4 UKBB hearing questionnaire responses over time

During the repeat visit conducted at Cheadle, ~20,000 participants responded to the hearing-related questionnaires for a second time (T1), 2-4 years after their baseline assessment, T0 (as described in 3.2.1.2 for the participants that repeated the SIN test). For the HD question, 80.21% of participants (N=15,940 of N=19,873, Table 3.13) submitted the same response, either 'Yes' or 'No' to the question at both T1 and T0. 9,2% responded 'No' at T0 and 'Yes' at T1, indicative of a hearing deterioration, while 3.43% responded 'Yes' at T0 and 'No' at T1, indicative of an improved hearing ability. The latter may be due to a temporary hearing loss, uncertainty of own symptoms or an error in response.

Similarly, for the HDBN question, 79% submitted the same response at both of the two timepoints while 10.67% responses indicated a deterioration between T0 and T1, and 6.78% indicated an improvement in HDBN (or ambiguous responses) between T0 and T1. Lastly, 94.19% of the sample submitted the same response at T0 and T1 to the

hearing aid use question. 5.02% indicated that they had started using a hearing aid frequently since the baseline visit, while 0.63% indicated that they were no longer using a hearing aid most of the time, having previously responded that they were at T0. The summary totals that were used to derive these percentages are listed below in Table 3.13.

HD		T0	
		Yes	No
T1	Yes	4,292	681
	No	1,826	11,484

HDBN		T0	
		Yes	No
T1	Yes	6,380	2,168
	No	1,378	9,677

HAID		T0	
		Yes	No
T1	Yes	568	573
	No	72	10,183

Table 3.13. Table listing the number of "Yes" and "No" responses at T0 and T1.

HD, UKBB question 'Do you have any difficulty hearing'; HDBN, UKBB question **"**Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?"; HAID, UKBB question 'Do you wear a hearing aid most of the time?"; Yes, "Yes" response to relevant question'; No, "No" response to relevant question; T0, baseline visit; T1, first repeat visit at Cheadle assessment centre. All values in the table represent numbers of participants with the corresponding two responses at T0 and T1. Participants that selected any other response for either question are not included in the totals listed in this table.

3.2.2.5 A comparison between SRT scores and questionnaire measures

In addition to using the questionnaire measures as surrogate phenotypes for ARHI, the questions could be used to validate the SRT data and *vice versa*. Anomalies or false readings could be detected by comparing participants SRT scores with questionnaire responses. For all three responses, it is expected that the SRT_B mean and median for the subgroups that responded "Yes" would be greater than the mean and median for the subgroups that responded "No". Table 3.14 displays the mean, median and range of the SRT_B scores for Yes/No responses to each of the three hearing questions. The mean and median SRT_B scores were higher for the "Yes" respondents than the "No" respondents for all three questions.

	SRT_B mean	SRT_B median	SRT_B range	N
HD Yes	-6.84	-7	-11.25 8.00	47,570
HD No	-7.48	-7.5	-11.5 8.00	123,238
HDBN Yes	-6.90	-7.5	-11.25 8.00	65,994
HDBN No	-7.49	-7.5	-11.5 8.00	110,401
Hearing aid Yes	-5.44	-6	-11 8	5,437
Hearing aid No	-7.36	-7.5	-11.5 8.0	17,5381

Table 3.14 A comparison between SRT_B scores and questionnaire responses.

HD Yes; "Yes" response to the question 'Do you have any difficulty hearing', HD No; "No" response to the question 'Do you have any difficultly with your hearing?', HDBN Yes; "Yes" response to the question "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" HDBN No; "No" response to the question "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?", Hearing aid Yes; "Yes" response to 'Do you wear a hearing aid most of the time?", Hearing aid No; "No" response to 'Do you wear a hearing aid most of the time".

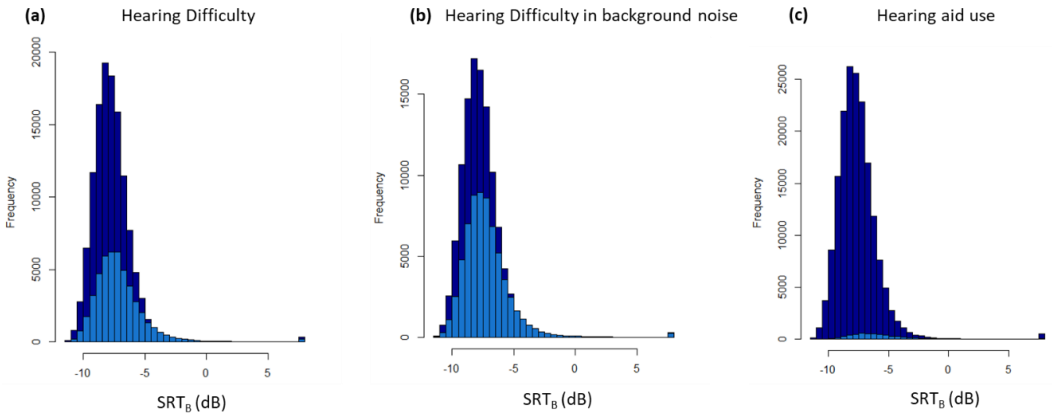

Figure 3.15. Histograms of SRT_B scores coloured by "Yes" and "No" responses to questionnaire measures.

The distribution of (a) and subsequent colouring corresponds to the SRT_B score of each participant and the response to the question 'Do you have any difficulty with your hearing?'. The distributions and colouring of (b) and (c) correspond to the responses to the questions 'Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing?)' and 'Do you wear a hearing aid most of the time?'. Light blue represents the distribution of SRT scores for "Yes" responses to the three questions and dark blue represents the distribution of SRT scores for "No" responses to the three questions.

Figure 3.15 displays three histograms of SRT_B scores. The histograms are coloured by participant response to each of the three UKBB hearing questions. Visually, the SRT_B scores coloured in light blue are slightly shifted to the right for both HD and HDBN, consistent with the mean scores of "No" respondents being higher than "Yes" respondents. As the proportion of "Yes" responses is much smaller in the Hearing aid responses sample, it is not clear from the visualisation, but the mean score is almost 2dB > for "Yes" respondents than "No" respondents (Table 3.14). Mann Whitney tests confirmed that the SRT_B distributions are significantly different between "Yes" and "No" respondents for each question (p-value < 2.2e-16). This is evidence that overall in the

sample, on some level, the SRT scores reflect personal perception of hearing ability and also of common hearing aid use.

The responses were also compared to the categorised UKBB scores; 'normal hearing' SRT < -5.5dB, 'insufficient' hearing SRT -5.5dB to -3.5dB and 'poor' hearing SRT > -3.5dB, as displayed in Table 3.15. The number of individuals with each classification (good, insufficient, poor), according to their SRT_B score, is listed against the questionnaire responses. The percentage of each category for the responses is also noted. For all questions, the percent of individuals with 'poor' and 'insufficient' hearing was greater in the 'Yes' category, as expected.

Question	Resp.	Normal	% of response	Insufficient	% of response	Poor	% of response	N total
HD	Yes	37,723	85.48	4,318	9.78	2,089	4.73	44,130
HD	No	111,096	94.26	5,488	4.66	1,279	1.09	117,863
HDBN	Yes	54,071	87.7	5,256	8.52	2,327	3.77	61,654
HDBN	No	99,622	94.31	4,893	4.6	1,116	1.06	105,631
Hearing aid	Yes	3,031	63	956	19.6	827	17.2	4,814
Hearing aid	No	154,547	92.74	9,443	5.67	2,660	1.60	166,650

Table 3.15. Comparison of SRT_B scores and questionnaire responses.

This table summarises the proportion of respondents to each question subset by SRT_B score when classified 'Good', 'Insufficient' and 'Poor'. Resp. Response to question; HD, 'Do you have any difficulty with you hearing?'; HDBN, 'Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing?)' and 'Do you wear a hearing aid most of the time?' and Hearing aid; 'Do you use a hearing aid most of the time?'

Perhaps surprisingly, 63% of participants that responded "Yes", they do wear a "hearing aid most of the time" have 'Normal' hearing. Similarly, >54,000 individuals that respond "Yes" they do have difficulty hearing in background noise (87.7% of "Yes" respondents), have 'normal' hearing as determined by the SIN. These figures indicate that either the SIN test is not as sensitive as other SIN tests, as explored previously and, or that these SIN classifications are not an accurate description of hearing ability for this specific SIN test, as previously proposed. Or, alternatively, that individuals over-report hearing difficulties (though this assumingly does not apply to the response to hearing aid use).

3.2.3 Phenotypes derived for association analysis: *HDiff* and *HAid*

It has been established in the previous sections that the combination of two "Yes" responses to the two questions regarding hearing difficulty results in a prevalence similar to the expected prevalence of ARHI in a sample of this age range; 21.62%. There is also a greater proportion of males reporting a hearing difficulty, as expected for ARHI. Findings from the SIN test data indicate that a number of limitations ought to be addressed prior to using the data as a surrogate ARHI phenotype in association analysis. In addition, the overall size of the sample that responded to the questions is three times the size of the sample that completed the SIN test, and so this data may provide greater power for association analysis.

Therefore, for association analysis, the two questions regarding hearing difficulty were used to derive a qualitative phenotype. Participants were assigned case/control status based on their responses to the questions. A second phenotype was also derived using the responses to the question regarding hearing aid use. This derived phenotype depicts hearing loss that has been clinically diagnosed and for which treatment is likely resulting in some relief of symptoms.

The first phenotype was derived from the two questions regarding hearing difficulty "Do you have any difficulty with your hearing?" and "Do you have any difficulty with your hearing" and "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" Participants that responded "Yes" to both questions were assigned as 'cases' while those that responded "No" to both of the two questions were assigned as 'controls.'

Participants with any other combination of responses were removed. In addition, controls aged <50 were removed from analysis in order to better age-match the case and control groups prior to analysis. To reduce the chance of false-positive controls due to an ambiguous combination of responses to the questions, controls that responded "Yes" to the question "Do you use a hearing aid most of the time?" were also removed. The resulting phenotype and will here-on be referred to as *HDiff* (Figure 3.17).

The responses to the question regarding hearing aid use were used to derive the second trait representative of hearing impairment. Participants that responded "Yes" to the question "Do you use a hearing aid most of the time?" were classified as 'cases' and those that responded "No" were defined as 'controls'. This phenotype will here-on be referred to as *HAid* (Figure 3.16). For both derived phenotypes (*HDiff* and *HAid*), if participants answered the questionnaire twice, i.e. attended an assessment centre for a repeat visit,

the answer at the second time point was used in analysis, in order to increase the mean age of the sample. The likelihood of including congenital forms of deafness was reduced as all participants that had selected 'I am completely deaf' in the UKBB questionnaire were excluded from analysis (based on the findings in Section 3.2.2.3)

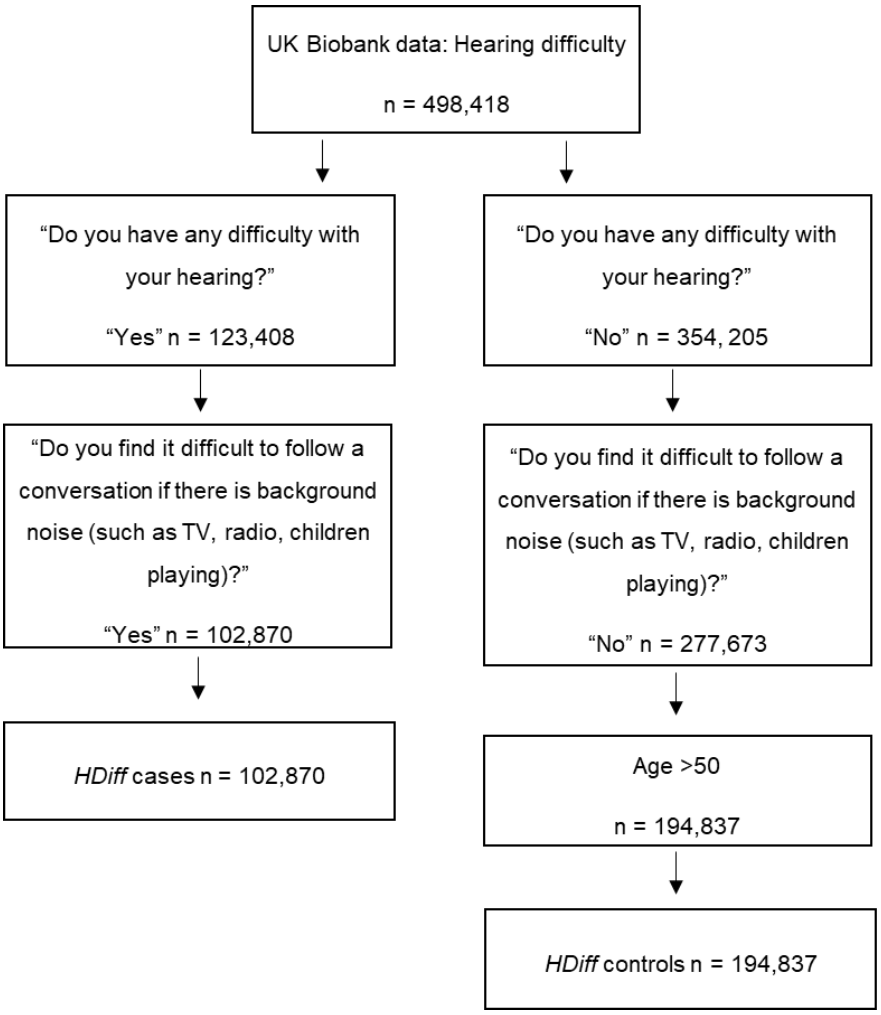

Figure 3.16. Flowchart describing how the phenotype based on self-report hearing difficulty was derived. n, number of participants.

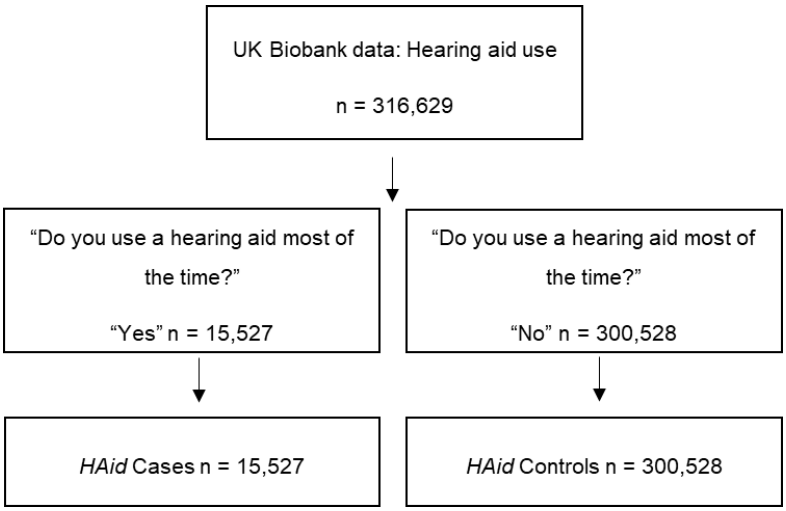

Figure 3.17. Flowchart describing how a phenotype based on hearing aid use was derived. n, number of participants.

3.3 Discussion

The aim of this chapter was to identify or derive a surrogate phenotype for ARHI, using data from the UKBB sample. The SIN test data and hearing-related questionnaire responses were assessed as to whether the data (1) conforms to established ARHI characteristics; symptoms of hearing loss, especially in speech intelligibility, that show a positive correlation with age, increased prevalence in male subjects, and a bilateral deterioration and (2) their practicality for use in genetic association analysis; a stable, unambiguous qualitative or normalised quantitative measure.

Firstly, the SRT scores were assessed to determine whether they are suitable as a quantitative phenotype for association analysis. The SRT distributions (Figure 1) are highly skewed (>1) to the right, a characteristic of SIN tests[101]. As a potential quantitative phenotype for association analysis however, this is a limitation; a normal distribution is one of the assumptions of regression models used in association analysis[274]. Although this is a large sample, a strong transformation is required which would result in a loss of sensitivity and thus a loss of power in the association analysis.

An alternative way of using the SRT scores as a phenotype for association analysis is to create a qualitative phenotype where cases and controls are assigned by pre-defined SRT thresholds that define 'poor/insufficient' and 'normal' hearing. However, when these pre-defined classification thresholds are applied, the prevalence of hearing impairment (poor or insufficient) in the sample is <12%. While these thresholds are defined by the laboratory standard of 'normal' scores (being within 2 standard deviations from the mean[264], here the mean of the normative sample), the subsequent prevalence is far lower than expected for ARHI in a cohort of this age range (40-69 year olds). Previous estimates predict that over 40% of the population aged 50 have hearing loss[96].

ARHI is understood to increase in prevalence with increasing age, is more prevalent and has an earlier onset in male subjects and presents as a bilateral hearing loss[85–88]. A suitable surrogate measure of ARHI would exhibit these three characteristics in the study sample. The UKBB SRT_B does not however conform to these characteristics. Firstly, the SRT_B correlation with age is $r^2_s = 0.238$, $p<2.2e-16$, indicative of a positive correlation with age but one that is relatively low for an age-related trait. In the UKBB sample SRT_B scores are not significantly different between males and females ($p= 0.065$). This supports a previous report that the male UKBB sample does not have an increased risk of 'poor' hearing (SRT_B > -3.5dB) and only had a slightly reduced risk of 'insufficient' hearing (-3.5dB < SRT_B > -5.5dB)[123]. Conversely, this conflicts with reports of previous

measures of ARHI, including SIN tests, that display an increased prevalence and severity among male participants[85–88].

There are multiple possible explanations for these findings. Firstly, it could be speculated that these classification of 'poor' and 'insufficient' may not represent ARHI but define a more severe and less prevalent form of hearing loss. Secondly, an alternative explanation, is that the elevated male prevalence in other samples and from the self-report measures, are due to elevated awareness rather than true hearing ability; this is the largest data sample that has not been selected based on perceived hearing ability. Furthermore, some cohorts that have been assessed via PTA testing are susceptible to a recruitment bias based on perceived hearing ability. A third explanation could be that the prevalence of genuine hearing impairment in the UKBB is slightly lower than the general population (and thus of previous estimates), due to its "healthy volunteer" selection bias that has been explored and reported elsewhere[275].

Analysis also challenges the understanding that ARHI generally presents as a bilateral hearing deterioration. The SRT_L and SRT_R correlation is 0.358, p-value< 2.2e-16 and the range of difference is 0-18dB. This either indicates that the prevalence of pronounced unilateral hearing deterioration is greater than is currently understood, or that a proportion of participants completed the test inconsistently due to modifiable external factors.

The results presented in this chapter demonstrate that SIN test data is currently not suitable as either a qualitative or quantitative surrogate phenotype measure for ARHI. SRT scores in this cohort do not conform to expected characteristics of ARHI. In addition, the SRT distribution would require a strong transformation to achieve a normal distribution and the SRT thresholds for hearing impairment would need to be re-defined to accurately represent ARHI in the sample. This work has identified limitations to the SRT data that ought to be addressed prior to any further work being conducted on this data.

In line with the findings presented here, a number of studies have also reported that this SIN test may be a less sensitive than alternative SIN tests[122,162,189]. The insensitivity of the test is apparent in the UKBB sample by the negligible difference between the interquartile range of the SRT_B distribution (2dB) and the mean difference between participant SRT_L and SRT_R (1.55dB). These two measures demonstrate the redundancy of the SIN measure in this sample; the test does not discriminate between the hearing abilities for the majority of individuals.

The UKBB-specific adaptations that were made to the SIN test could be where the relative insensitivity arose. The two main adaptations were the number of trials presented to an individual, and the number of trial ratio scores used to calculate the SRT. Both were reduced in the UKBB version of the test compared to previous tests. In theory, reducing these can make the test more efficient as it takes less time for the participants to complete the test, yet this may in turn reduce the sensitivity of the test.

Work that has been conducted since the implementation of the UKBB SIN test, has demonstrated that improving the test efficiency (reduce the number of trials from 25 to 8.3) can be achieved, but when using a fixed-SNR procedure[276] rather than the adaptive procedure (as used in the UKBB SIN test). The nature of the test procedure, fixed or adaptive, must be set prior to carrying out the test. Therefore, it is not a feasible retrospective alternation to make to the UKBB data.

The number of trials completed by each participant must also be set prior to commencing the test. The number of trials used to calculate the SRT score can however be altered following the test; the maximum amount being the number of trials the participant completed. In this case, SNR scores from 8 trials were used to calculate the SRT, but SNR scores from all 15 trials were stored. Using all 15 SNR values to calculate the SRT may increase the sensitivity or reliability of the SRT score, as a greater number of data points will be used to calculate the mean value (the resulting SRT_L or SRT_R scores). However, the UKBB sample has an additional, unique, limitation due to the protocol change. Either the sample would need to be split into two distinct groups for analysis, or the SRT value adjusted according to which protocol was used.

Furthermore, the differences in SRT distributions between assessment centres and the difference in SRT measures over time in the same individuals, is suggestive of an unstable measure. A previous study reported a small numeric change in SRT scores over five years for different age groups (0.27dB for 40-49-year olds, 0.51dB for 50-59-year olds and 0.41dB for 60-69-year olds)[272]. Another study that conducted longitudinal analysis for SIN testing and reported a test re-test correlation of 0.7 with a mean score difference of 0.229[277]. The test re-test correlation was lower in the UKBB sample; 0.25 for SRT_L and 0.23 for SRT_R.

There are however clear limitations when making direct comparisons between these different datasets such as the age range, format of SIN tests and time elapsed between T0 and T1. Although the time period between tests was relatively short in the case of the UKBB test, further sources of variation could be present such as environmental factors

or even the progression of associated traits that affect the progression of a HL at different rates between individuals. A sensitive and stable measure would have a proportionally greater interquartile range (less redundant) yet a greater consistency within participants (greater test re-test correlation), respectively.

The power to detect genetic variants that are associated with a trait of interest is dependent on the amount of the variation in the trait that can be explained by genetic factors. Therefore, minimising non-genetic effects is required to maximise power. Potential modifiable environmental factors that affect the SRT score could be identified and quantified, and the SRT scores either adjusted or excluded accordingly. Speculatively, modifiable environmental effects between the assessment centres may be the differing noise levels present in the assessment centres, the frequency of headphone calibration, the maintenance of systems and additional instructions that are relayed to participants by assessment centre staff.

In addition to the possible differences between the assessment centres themselves, there may be geographical differences in SRT scores that are due to lifestyle factors such as the prevalence of occupational noise-exposure or even the interpretation of the SIN test speech due to local accents. Previous work has even demonstrated that factors such as participant computer literacy and education has an effect on the SRT_B test score[160]. Here, the effect of individual assessment centres or participant computer literacy on the SRT score could be quantified and the data adjusted accordingly. By modelling the effect of each of these potential factors on the SRT score, it is possible to quantify the proportion of variance within the SRT distribution that is caused by each potential factor.

Using this method, it may be possible to stratify participants based on assessment centre, protocol used or testing time period and use a meta-analysis approach. However, there would likely to be an element of collinearity which could result in an over-adjustment; computer literacy likely has a relationship with age and education. Similarly, demographics vary between assessment centres and so the varied SRT distributions may be a true reflection of hearing abilities rather than simply due to different testing practices at between assessment centres.

To further investigate the SRT data, we shared our findings with Cas Smits, author of the original digits in noise test[188]. We visited the Smits lab in the Netherlands and presented the findings from the UKBB SRT data that is presented here in this chapter. Data from this original speech in noise test is used extensively in this chapter, in section 3.2 to compare to the data trends observed with the UKBB SRT data. Cas Smits was able to

identify further discrepancies between SRT score distributions between UKBB assessment centres, such that the scores are not directly comparable between centres. This finding supports the finding that SRT distributions were found to be significantly different between different UKBB assessment centres. Cas Smits subsequently published a study on the effects of a reduced number of triplets used in speech in noise tests and the resulting reduction in test sensitivity when an adaptive procedure is followed[276]. This is a possible factor contributing to the limitations identified in this chapter and could be further investigated with the UKBB data specifically. The findings presented in this thesis and subsequent concerns regarding the viability of the speech in noise data were also communicated with UKBB directly.

In order to use the SRT data in future analysis, there are a number of ways that these limitations could be addressed. One way to quality-check the SRT data would be to cross-reference the SRT scores with participant responses to the questionnaire measures. As presented in section 3.2.2.5, the questionnaire measures and SRT scores do have some level of consistency when comparing the mean and median SRT_B values between "Yes" and "No" respondents to each of the questions. However, as can be deduced from the data in Table 3.15 where questionnaire responses are compared with the classifications of 'normal', 'insufficient' and 'poor' hearing, the data is conflicting. 63% of participants that claim to wear a hearing aid most of the time and 87% of respondents that have difficulty hearing in noise have 'normal' hearing according to their SRT_B score classifications.

The questionnaire data could even be used to re-define the classifications based on the prevalence of HL reported in the sample and the population. The authenticity of +ve dB SRT scores, especially +8dB SRT scores could also be checked via this method. Participants that reported 'I am completely deaf' did not take part in the SIN test, and therefore all participants that took part are expected to have some ability to hear in at least one ear. Therefore, a score of +8 may indicate that an individual did not perform the test correctly due to a misunderstanding or a malfunction of equipment. To reduce the likelihood of false +8dB scores, individuals that scored +dB for both ears and/or that did not report any hearing difficulties on the questionnaire, ought to be removed from further analysis.

Following on from the SIN test analysis, the UKBB questionnaires and corresponding results were assessed for use as ARHI-surrogate phenotypes for genetic association analysis. Categorising participants based on "Yes" and "No" responses from the questions regarding hearing difficulty and hearing aid frequency, results in a qualitative

phenotype. “Yes” responses determine ‘cases’ and “No” determine ‘controls’. The questions regarding hearing difficulty resemble suitable surrogate measures for ARHI; reported hearing difficulty increases with age and a higher prevalence is observed in male subjects. The combination of two “Yes” responses to the hearing difficulty questions is the most accurate indication of ARHI in this sample based on the prevalence that is observed and increased proportion of males. The two hearing difficulty questions capture the main symptoms of ARHI; difficulty hearing and specifically, a difficulty hearing in the presence of background noise. The prevalence of the hearing difficulty measures and of hearing aid use are in line with both ARHI prevalence and hearing aid use for this age group, despite data demonstrating the under or over-estimation of hearing impairment by self-report measures[278].

Again, there are limitations that ought to be considered when using the questionnaire-derived phenotypes. The main limitation is the bias that results from the subjective nature of the questions and possible responses. This limitation is not restricted to hearing phenotypes, but to many traits; the way in which individuals experience and thus classify symptoms is subjective. For example, two individuals may clinically have the same level of hearing impairment, but one is more aware of or more distressed by the symptoms than another; this individual is likely to report a greater hearing impairment. Due to this, there can be a disparity between an individuals’ true hearing impairments and their perceived (and thus self-reported) hearing impairment. This is difficult to adjust for within a sample, but one way to reduce this effect is to limit the use of questions that elicit an emotional response by altering the language that is used. A second method is to reduce the ambiguity of responses.

The two questions regarding hearing difficulty are designed such that these effects are minimised. Firstly, the two questions directly ask, ‘do you have’ rather than ‘do you feel’ or ‘to what extent to do think you have’, and therefore evoke less of an ‘emotional’ response. Secondly, the possible answers are clearly defined; “Yes” and “No”, or “Prefer not to answer” or “Do not know”. This also reduces the effect of individuals’ self-perceptions on the classification; a question asking participants to rank their hearing as ‘very good/good/poor/’ would introduce further bias to the responses as there is more variation in individuals’ perceptions of ‘good’ vs. ‘very good’ than ‘yes’ and ‘no’. To further combat the issue of subjective responses to questionnaire measures, I have used a combination responses from more than one question to define the phenotype, which in turn increases the sensitivity and accuracy of the measure[88,279].

There is a limitation in the way that the UKBB hearing difficulty questions are phrased however; no information is captured relating to the age of onset. Considering this, as the minimum age of participants is 40 years, a proportion of the control group may have the same genetic risk for developing the trait as participants in the case group but are yet to develop symptoms. For this reason, it is appropriate to age-match the two groups for association analysis, in this sample this was achieved by removing all controls aged <50 years. In addition to the unknown age of onset, the cause of hearing impairment is also unknown. While the majority of the individuals that report a hearing difficulty are expected to be experiencing ARHI (as this is the most common form of hearing loss in the population), it is likely that there is a proportion of participants suffering from hearing loss by a separate cause. Again, using a combination of responses from both of the questions can reduce this effect.

Similar to the hearing difficulty question responses, the responses to the hearing aid use question are "Yes", "No", "Prefer not to answer" and "Do not know" which again avoid bias from a possible 'emotional' response. The question is however more subjective; participants are asked whether they wear their hearing aid 'most' of the time. The amount of time that two individuals wear their hearing aids per week could be exactly same, but their response to this question may differ simply by their own interpretation of 'most of the time'. Despite this limitation, this question is the closest measure in the UKBB to a clinical hearing loss diagnosis; frequent hearing aid uses will have been diagnosed with a hearing impairment and have subsequently been prescribed with a treatment device.

This question can therefore serve as a quality control measure for both the accuracy of the derived phenotypes and the results from the genetic association analysis. As described earlier, this measure could be used to validate SRT scores and identify false-positive cases or controls. In addition, the cases and controls defined with the hearing difficulty questions can be cross-checked with the responses to this question. Participants that answered "No" to the hearing difficulty questions and "Yes" to the hearing aid use question may have done so, if they do not have any difficulty with their hearing while using a hearing aid. In this study however, it is the participant's hearing ability without the use of hearing aids that is of interest. Therefore, this question can be used to remove any participants that were falsely assigned 'controls' based on this interpretation of the hearing difficulty questions.

The surrogate-ARHI phenotypes derived in this chapter (*HDiff* and *HAid*) display the hallmarks of ARHI that are possible to test; an increase in prevalence with increasing age

and in male subjects. However, as the possible responses are “Yes” and “No”, not further subjective classification is required to create a qualitative phenotype. The two qualitative were determined to be suitable for use in genetic association analysis. They also provide an opportunity to test the validity of phenotypes derived from questionnaire responses, in hearing genetic association analysis.

Chapter 4 - Genetic association analysis

4.1 Introduction

ARHI is understood to be a common complex trait, caused by multiple genetic and environmental risk factors. However, as discussed in chapter 1, little is understood about how these risk factors contribute to a pathology. GWAS is a method that can be used to identify trait-specific genetic risk variants, in the absence of any prior knowledge of trait-specific mechanisms or biological pathways. Such genetic risk variants can subsequently be used to predict the biological underpinnings of a trait[213]. This method therefore provides an important opportunity for ARHI research. The last decade of common complex trait genetic research has been termed the 'GWAS era' as the method has led to the discovery of thousands of loci associated with multiple common complex traits[219,280].

Prior to the advent of GWAS, linkage analysis and candidate gene analysis were commonly used to identify novel risk loci for common complex traits such as ARHI. In linkage analysis, family pedigrees are studied to identify regions of chromosomes that are inherited in conjunction with a phenotype of interest. Linkage analysis was commonly used for complex traits prior to the development of SNP arrays and high-density genotyping, as described in chapter 1 section 1.4.1. Several linkage studies have been used to identify genome regions implicated in hearing loss[206,207], yet the method is most optimal for detecting highly penetrant pathogenic variants causal of Mendelian disorders. This is because the method generally lacks the power required to discover multiple variants of small effect sizes that cause complex traits such as ARHI.

Candidate gene analysis is an alternative method used to test for association between a trait of interest and specific genome regions. It can be a more accessible approach than GWAS as it does not require genome-wide, high density genotyping, and so is less economically and computationally demanding. Candidate gene analysis is however dependant on prior knowledge of links to a genomic region or biological pathway of interest; a limitation for current ARHI studies.

As described previously in section 1.4, GWAS is a hypothesis-free method where SNPs across the genome are tested for association with a trait of interest in a population sample. Current GWAS methods are a product of the rapid advances that have been made to the methodologies employed to uncover genetic components of common complex traits. As the success of GWAS has grown for numerous complex traits, so has the size of study samples, range of phenotypes studied and the number and complexity of the computational methods. The optimal statistical method for association analysis depends

on the type of data that is used. Quantitative phenotypes are analysed by generalized mixed models, under various model assumptions. Qualitative traits are generally analysed with logistic regression analysis. Most commonly, an additive model is applied in GWAS, but an alternative model can be used, based on the assumption that there are alternative genetic effects are present such as dominant, recessive, multiplicative or additive SNP variants. During the GWAS era, models have been developed such that data with different structures and formats can be comparatively analysed.

As described in Chapter 1 section 1.4.1, ARHI is in the infancy of its GWAS era. The first five ARHI GWAS did not identify any genome-wide significant loci. All five used PTA frequency thresholds to derive phenotypes for the association analysis; the studies used either selected PTA frequency thresholds, PC analysis or z-scores derived from frequency thresholds. The first ARHI GWAS study used population samples pooled from 9 centres in 7 European countries, with a discovery sample totalling n=1692[109]. Although no genome-wide significant associations were found, authors reported rs11928865, in the *GRM7* gene transcript, as the most highly associated SNP and replicated at their Bonferroni corrected threshold.

The second study, on the Saami population (n=352), used PC analysis of PTA thresholds to define the phenotype and reported rs457717, in the intron of the *IQGAP2* gene transcript, as the most highly suggestive association[107]. The third study was another meta-analysis of European samples and highlighted suggestive associations with the *DCLK1, PTPRD, CIMP* and *GRM8* gene transcripts[114]. The fourth deployed a logistic regression analysis of PTA thresholds on a sample of 3,900 participants from the UK. Here the focus was on *ESRRG,* where an association was observed in females only[92]. In 2015 Fransen performed an ARHI GWAS with the aim of replicating these previous findings. The authors of this study hypothesised ARHI to be a highly polygenic trait, with multiple causal genetic variants that have relatively small effect sizes that are not possible to detect with 'modestly powered' GWAS[38].

The first genome-wide significant associations in an ARHI GWAS were identified in 2014. A SNP in an intron region of the *SIK3* gene transcript was identified using a phenotype derived from PTA thresholds by PC analysis, in a meta-analysis of European samples n=4939[111]. Later that year, a second study identified significant SNPs in close proximity to the *SLC28A3* and *PCDH20*[112] gene transcripts. These two associations were identified in a meta-analysis of isolated populations in Italy and Central Asia and were further replicated in British and Finnish samples. Most recently, in 2016, the largest ARHI GWAS to date was performed using a variety of phenotypes derived from data stored in

electronic health records in the US. The size of the study sample was a magnitude larger than any previous ARHI GWAS, with n >6500 ARHI cases and n >45,000 controls. Significant associations were identified within and in close proximity to *ISG20* and *TRIOBP* gene transcripts and the associations were further replicated in the independent UKBB sample[113].

Collectively, these studies reveal only a limited number of genetic variants to be significantly associated with ARHI via GWAS in population samples (see Table 1.5). The relatively few significant findings, combined with the multiple substantial heritability estimates, lead to a hypothesis that, like other complex traits, common adult hearing impairment is highly polygenic[38]. However, due to the recent availability of large-scale cohort data, namely the UKBB[245], and improved genotyping and imputation coverage, there is now the potential for a more comprehensive association analysis to test the theory that ARHI has a highly polygenic nature, and to discover trait-associated variants.

For many traits the GWAS era has permitted the identification of numerous trait-specific variant associations. The initial post-GWAS steps have been described in detail in Chapter 1, and commonly consist of (i) identifying the most the SNP(s) that best represent the association(s) at each region of association,[231] (ii) validating the associations via replication analysis and (iii) exploring the biological relevance of the association, in the context of the trait of interest. Firstly, methods can be deployed that test for multiple associations at each locus and determine each lead SNP. Although the genome coverage of genotyping arrays and imputation reference panels have greatly increased, as described in Chapter 1 on page 62, the causal SNP may not feature on the array or panel and thus not be present in the association analysis. Where the data is available, the region of interest can even be sequenced, and the true functional variant can be identified.

Secondly, it is custom to replicate the association analysis in an independent cohort to ensure that the findings are not study or cohort-specific and that the findings can be applied to a larger population. The replication analysis, where possible, ought to apply the same analytical method on the same genetic variants, using a comparable phenotype definition, in the independent sample[231]. Thirdly, the biological functions of associated SNPs can then be investigated either individually and, or collectively to hypothesise trait-specific pathological mechanisms. A large number of *in vitro* and *in vivo* methods can be used to explore this. Further methods such as genetic correlation analysis[281] can also be applied to draw hypotheses about how different traits are linked by common biological functions and pathologies, based on a shared genetic variance. Hypotheses could then

even be created with the aim to determine causality between traits, as discussed in Chapter 1, section 1.3.3.

In this chapter genetic association analysis is performed with the aim of identifying genetic variants that are associated with the two phenotypes termed '*HDiff*' and '*HAid*' that were derived from UKBB questionnaire measures (Chapter 3). Secondly, replication analysis is performed, along with calculations to estimate the SNP heritability estimates for *HDiff* and HAid the two phenotypes in the UKBB sample. Lastly, gene-set analysis is conducted with the aim of identifying enriched biological pathways and mechanisms, and genetic correlation analysis conducted with the aim of identifying traits that share variance based on common genetic variants.

4.2 Results

The results for this chapter are presented in three main sections as depicted below in Figure 4.1.

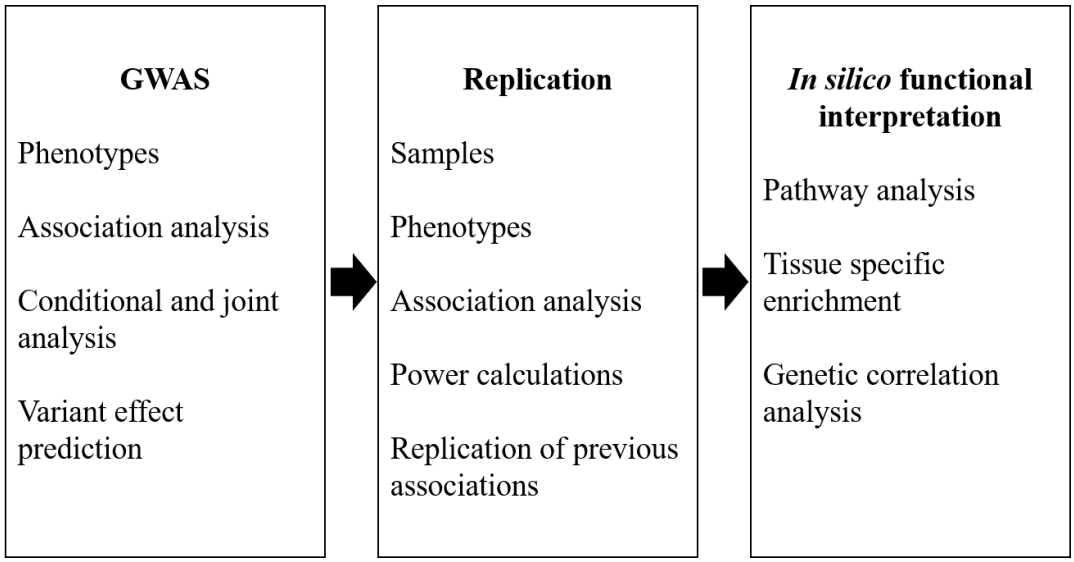

Figure 4.1. Diagram to illustrate the layout of the results section for chapter 4.

The three sections are (1) GWAS (2) Replication (3*) In silico* functional interpretation.

4.2.1 Genome-wide association study

4.2.1.1 UKBB phenotype definitions for GWAS

In Chapter 3 two phenotypes were derived for use in genetic association analysis. UKBB participants were categorised using a case-control design based on responses to questions regarding hearing difficulty (*HDiff*, n=498,281) and hearing aid use (*HAid*, n=316,629). The description of how these two phenotypes were derived is detailed in Chapter 3, section 3.2.3 and - Methods, section 2.1.1.2.

The samples were filtered based on genomic quality control criteria outlined in Chapter 2. This was in addition to the genomic quality control that was conducted centrally by UKBB. Individuals were removed based on high levels of heterozygosity (indicative of sample contamination), any samples with excess relatedness (samples which have more than 10 putative third-degree relatives in the kinship table) and selected based on 'white British' status (samples who self-reported 'White British' and who have the corresponding genetic ancestry based on a PC analysis of the genotypes). Following the quality control filters, the final sample sizes used for the two association analyses are n=250,389 for *HDiff* (87,056 cases and 163,333 controls, Figure 4.2), n=253,918 for *HAid* (13,178 cases and 240,740 controls, Figure 4.3).

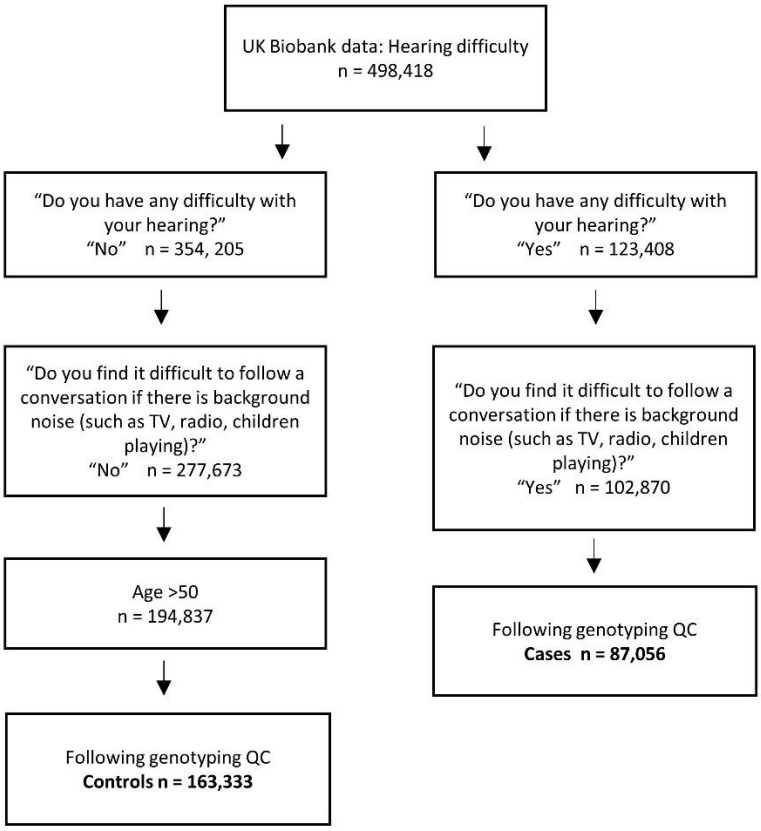

Figure 4.2. Flow chart describing case-control assignment for the hearing difficulty (*HDiff*) phenotype.

Participants answered questions as part of the UKBB questionnaire that was administered at UKBB assessment centres. Participants who responded, 'Prefer not to answer', 'I am completely deaf' and 'Do not know' were removed from the analysis. Participants were removed from the control group if they responded "Yes" to "Do you use a hearing aid most of the time?" A lower age limit of 50 was implemented for controls to ensure age was consistent between the case and control groups due to the association of ageing with the trait.

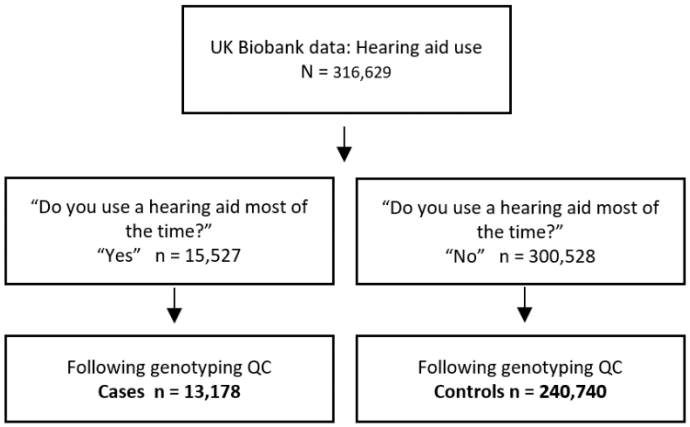

Figure 4.3. Flow chart describing case-control assignment for the hearing aid use (HAid) phenotype.

Participants answered questions as part of the UKBB questionnaire that was administered at the UKBB assessment centres. No information was collected regarding age at hearing aid prescription or cause of hearing loss.

4.2.1.2 Genetic association analysis

Two separate genetic association analyses were then conducted for *HDiff* and *HAid,* as detailed in - Methods, section 2.1.1.4 - 2.1.1.5. For both analyses, a linear mixed-effects model was used to test for association between the two phenotypes and 9,740,198 SNPs, using the analysis tool BOLT-LMM v.2.3.2[226]. The BOLT-LMM linear mixed effects model accounts for population stratification and cryptic relatedness present in large samples such as the UKBB, and jointly models all genotype markers. The two association analyses resulted in the identification of 2,080 SNPs associated with *HDiff* at genome-wide significance (P<5E-08) and 240 SNPs associated with *HAid* at genome-wide significance (P<5E-08), Figure 4.4.

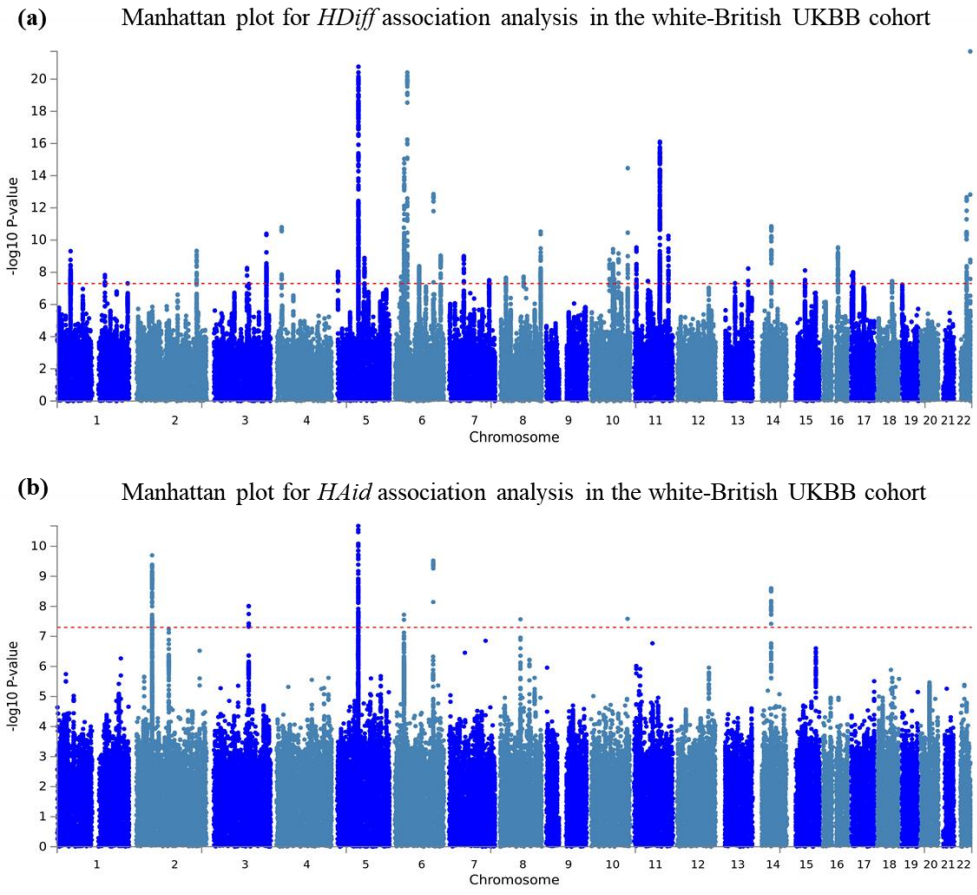

Figure 4.4. Manhattan plots displaying GWAS results for (a) *HDiff*, and (b) *HAid* phenotypes.

The Manhattan plots display the P values of all SNPs tested in discovery analysis. The threshold for genome wide significance (p<5E-08) is indicated by a red dotted line.

In certain datasets, it is common practice to assess the genomic inflation factor (λ_{GC}) to determine the amount of genomic inflation present and even to adjust the test statistics accordingly. However, when using large-scale datasets to analyse polygenic traits, this

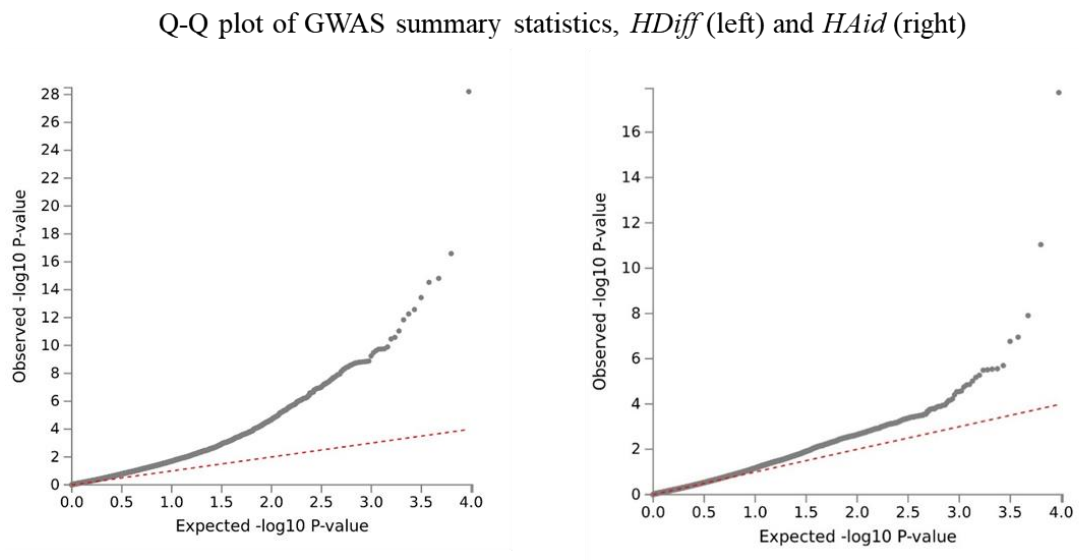

Figure 4.5. Q-Q plot of GWAS summary statistics, *HDiff* (left) and *HAid* (right).

The plots display the quantile distribution of observed p-values (y-axis) vs the quantile distribution of expected p-values (x-axis).

can be an overly conservative approach. Therefore, an approach such as LDSC can be used to estimate the proportion of inflation in the χ^2 statistic that the intercept attributes to alternative explanations than polygenicity, such as population stratification. Here, univariate linkage disequilibrium score regression (LDSC), (detailed in - Methods 2.2.1), was used to analyse the relationship between test statistic and LD[251].

Figure 4.5 displays quintile-quintile (Q-Q) plots for both *HDiff* and *HAid* analysis. The LD score regression intercepts for the two analyses are 1.032 for *HDiff* and 1.03 for *HAid*. The ratio (intercept-1)/(mean(χ^2)-1) for *HDiff* was 8% and 5% for *HAid*. These represent the proportions of the inflation in the χ^2 statistic that the intercept attributes to alternative explanations than polygenicity.

4.2.1.3 Conditional and Joint analysis

SNPs included in the analysis which are in proximity with the causal SNP at an associated locus, will likely have inflated test statistics due to LD. Therefore, not all significant SNPs represent individual associations, a cluster of SNPs in LD is likely representing only one or two causal variants. Rather than assuming that the SNP with the smallest P-value (and thus has the highest association) in each region of LD represents and captures the true effect of the variation at the locus, a conditional and joint analysis was conducted. The method tests for the presence of multiple causal variants at individual loci and whether the variation is accurately represented by a single SNP[232].

Here conditional and joint analysis using GCTA-COJO[232], as described in - Methods 2.2.2, estimated that the 2,080 and 240 SNPs represent 41 and 7 independent loci to be associated with *HDiff* and *HAid* respectively, and identified a lead SNP at each locus. The method calculates a joint p-value for the region (pJ-value, Table 4.1, Table 4.2). Where this value is greater than the lead SNP p-value, this indicates that the true variance at that region is underestimated by the effect of the single SNP.

4.2.1.4 Predicting the functional effects of variants using *in-silico* analysis

Following the identification of the 48 lead SNPs, the Variant Effect Predictor (VEP)[258] was used to establish the genomic context of each lead SNP. Each of the 48 SNPs was mapped to the nearest protein coding gene using the GRCh37 genomic reference. 27 of the *HDiff* lead SNPs and all seven of the *HAid* lead SNPs are located within gene transcripts. Of 41 lead SNPs associated with *HDiff,* six variants lie in exons, four of which are predicted to result in missense mutations in the following genes transcripts: *EYA4, CDH23, KLHDC7B* and *TRIOBP*. 21 of the remaining *HDiff* lead SNPs lie within introns and the 14 remaining *HDiff* lead SNPs are intergenic (Table 4.1.1). Six of the lead SNPs associated with *HAid* reside in intronic regions and 1 lead SNP is intergenic. In Table 4.1and Table 4.2, the nearest protein coding gene to each SNP is named, along with the distance from the lead SNP to the gene transcript.

Four gene loci feature lead SNPs in both *HDiff* and *HAid*: *ARHGEF28*, *NID2, CTBP2* and *EYA4*. Furthermore, two highly significant independent associations with *HDiff* are within 100 kb of the *ARHGEF28* gene (rs6453022, *p*-value=2.07E-12 and rs6890164, *p*-value=4.15E-10), and the locus is also highly associated with the *HAid* phenotype (rs4597943, *p*-value=2.10E-11). The names of the genes in close proximity to the lead SNPs, (Table 4.1 column 10 and Table 4.2 column 10) are annotated '**a**' where the gene has previously been linked to a hearing phenotype or has a known function in the auditory system. Ten genes, a quarter of those listed, have previous links to hearing phenotypes or auditory function and are discussed in detail in Chapter 5.

***HDiff* results**

Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*p*J-value	Genes within 100kb	Dist. To nearest gene (bp)
22	rs36062310	G	0.96	1.000	-0.0315	0.003	1.90E-22	1.92E-22	***KLHDC7B,*** *SYCE3, ADM2, ARSA[a], CHKB, CPT1B, LMF2, MAPK8IP2, MIOX, NCAPH2, ODF3B, SBF1, SCO2, SYCE3, TYMP*	0
5	rs6453022	C	0.50	1.000	-0.0126	0.001	1.70E-21	2.07E-12	*ARHGEF28*	0
6	rs759016271	AGTAGTCCACTTTTC TTCTTTGCCTG	0.39	0.997	-0.0127	0.001	6.10E-21	6.16E-21	***ZNF318,*** *CRIP3, SLC22A7, CUL9, DNPH1, TTBK1*	0
5	rs6890164	A	0.51	0.993	0.0119	0.001	3.30E-19	4.15E-10	***ARHGEF28***	6177
11	rs7951935	G	0.62	0.996	-0.0114	0.001	7.80E-17	7.85E-17	***TYR****, NOX4*	1472
6	rs35186928	G	0.62	0.991	-0.0109	0.001	1.70E-15	1.69E-15	***HLA-DQA1****, HLA-DRB1, HLA-DRB3, HLA-DRB5, HLA-DRB6*	13352
6	rs9493627	G	0.68	1.000	-0.0104	0.001	1.40E-13	1.41E-13	***EYA4[a]***	0
22	rs132929	G	0.59	0.999	-0.0098	0.001	2.20E-13	4.61E-13	***BAIAP2L2[a]****, SLC16A8, PICK1, PLA2G6, POLR2F*	0
22	rs5756795	T	0.54	1.000	-0.0092	0.001	5.10E-12	1.09E-11	***TRIOBP[a]****, GALR3, GCAT, GGA1[a], H1F0, LGALS1, NOL12, PDXP, SH3BP1*	0
14	rs1566129	T	0.41	1.000	0.0091	0.001	1.40E-11	1.37E-11	***NID2****, GNG2, RTRAF*	0
4	rs35414371	T	0.87	0.998	-0.0131	0.002	1.60E-11	1.64E-11	***CLRN2[a]****, LAP3, MED28, QDPR*	1965
3	3:182069497_ TA_T	TA	0.84	0.989	-0.0118	0.002	4.10E-11	4.07E-11	***ATP11B***	441791
11	rs12225399	G	0.65	0.989	-0.009	0.001	8.60E-11	8.67E-11	***PHLDB1****, ARCN1, IFT46, KMT2A, TMEM25, TREH, TTC36*	0
11	rs55635402	A	0.81	0.996	0.0105	0.002	2.90E-10	2.94E-10	***TUB[a]****, EIF3F, NLRP10, OR10A3, RIC3*	0
16	rs62033400	A	0.61	0.999	0.0085	0.001	2.90E-10	2.95E-10	***FTO****, RPGRIP1L*	0
8	rs13277721	G	0.49	0.992	-0.0083	0.001	3.30E-10	3.35E-10	***AGO2,*** *PTK2*	0

***HDiff* results**

Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*p*J-value	Genes within 100kb	Dist. To nearest gene (bp)
2	rs62188635	C	0.45	0.988	0.0083	0.001	4.70E-10	4.72E-10	***KLF7***	50519
6	rs2236401	C	0.49	0.997	-0.0081	0.001	9.30E-10	9.38E-10	***SYNJ2***[a], *SERAC1*[a], *GTF2H5*	0
7	rs4947828	T	0.23	0.999	-0.0096	0.002	1.00E-09	1.02E-09	***GRB10***	0
10	rs6597883	T	0.84	0.989	0.0111	0.002	1.00E-09	1.05E-09	***CTBP2***	0
5	rs34442808	T	0.49	0.992	-0.008	0.001	1.30E-09	1.32E-09	***MCTP1, SLF1***	0
10	rs835267	A	0.53	0.996	0.008	0.001	1.60E-09	1.58E-09	***EXOC6***, *CYP26A1, CYP26C1*	0
10	rs4948502	T	0.57	0.995	0.0081	0.001	1.70E-09	5.63E-10	***ARID5B***	0
10	rs10824108	G	0.42	0.999	-0.0079	0.001	3.00E-09	1.24E-08	***ADK***, *AP3M1, VCL*	0
1	rs12027345	G	0.57	0.995	0.0079	0.001	3.60E-09	3.64E-09	***MAST2,*** *GPBP1L1, MAST2, TMEM69, TMA16P2, GPBP1L1*	12668
6	rs217289	G	0.56	0.992	-0.0078	0.001	4.90E-09	4.92E-09	***SNAP91***	0
3	rs13093972	A	0.55	0.992	-0.0078	0.001	5.50E-09	5.56E-09	***ZBTB20***	121137
15	rs62015206	C	0.41	1.000	-0.0078	0.001	7.70E-09	7.76E-09	***MAPK6,*** *BCL2L10, GNB5*	15613
5	rs10475169	A	0.88	1.000	-0.0117	0.002	9.30E-09	9.37E-09	***IRX2***	190445
17	rs17671352	T	0.38	0.999	0.0078	0.001	1.00E-08	1.43E-08	***ACADVL***, *DVL2*[a], *DLG4, ASGR1, CLDN7, CTDNEP1, EIF5A, ELP5, GABARAP, GPS2, NEURL4, PHF23, SLC2A4, YBX2*	0
1	rs7525101	C	0.56	1.000	-0.0075	0.001	1.50E-08	1.45E-08	***LMX1A***[a]	61973
17	rs12938775	G	0.50	1.000	0.0075	0.001	1.60E-08	2.25E-08	***PAFAH1B1***, *CLUH, RAP1GAP2*	0
8	rs76837345	A	0.93	0.997	-0.0146	0.003	1.90E-08	1.95E-08	***CHMP4C***, *IMPA1, SLC10A5, SNX16, ZFAND1*	0
6	rs9366417	G	0.26	0.993	0.0085	0.002	2.10E-08	2.12E-08	***SOX4***	291019
8	rs3890736	G	0.63	0.993	-0.0077	0.001	2.20E-08	2.22E-08	***GFRA2***	15676

***HDiff* results**

Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*p*J-value	Genes within 100kb	Dist. To nearest gene (bp)
10	rs143282422	G	0.99	1.000	-0.0349	0.006	2.40E-08	3.02E-08	***CDH23***[a], *C10orf105*	0
7	rs9691831	A	0.42	0.995	-0.0074	0.001	3.10E-08	3.11E-08	***TMEM213***, *ATP6V0A4*[a], *KIAA1549*	0
11	rs141403654	A	0.98	0.878	-0.0313	0.006	3.50E-08	3.53E-08	***AGBL2***, *C1QTNF4, FNBP4, MTCH2, NUP160*	0
18	rs4611552	T	0.78	0.995	-0.0089	0.002	3.60E-08	3.56E-08	***CCDC68***	9362
13	rs12552	A	0.44	0.994	0.0073	0.001	4.80E-08	4.86E-08	***OLFM4***	0
1	rs10927035	C	0.35	0.995	-0.0075	0.001	4.90E-08	4.89E-08	***ATK3***, *SDCCAG8*	0

Table 4.1. Results output from BOLT-LMM and GCTA-COJO for *HDiff* association analysis.

The table is continued on from the previous pages. Results output from BOLT-LMM and GCTA-COJO. Chr., chromosome; SNP, single nucleotide polymorphism; EA, effect allele; EAF, frequency of effect allele in BOLT-LMM; INFO, imputation quality metric, combination of imputation score and dosage confidence; β, effect size from BOLT-LMM approximation to infinitesimal mixed model; SE, standard error of the effect size; p-value, infinitesimal mixed-effects model association test p-value; pJ-value, p-value from a joint analysis of all the selected SNPs; Genes within 100kb, protein-coding genes that are within 100kb of the lead SNP. The gene in closest proximity to the SNP is emboldened; Dist. to gene (bp), distance in base pairs between SNP and nearest gene, a value of 0 indicates the SNP lies within the gene. a denotes genes previously linked to hearing phenotypes in mice or humans. Two SNPs reached genome-wide significance that are in close proximity to HLA-DQA1 on Chr 6 (Figure 4.4) but were not present in conditional analysis results.

***HAid* results**

Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*p*J-value	Genes within 100kb	Dist. to nearest gene (bp)
5	rs4597943	G	0.51	0.989	-0.0042	0.001	2.10E-11	2.09E-11	***ARHGEF28***	0
2	rs9677089	A	0.75	0.989	-0.0046	0.001	2.00E-10	1.98E-10	***SPTBN1***[a]	0
6	rs9321402	G	0.68	0.999	-0.0042	0.001	3.00E-10	3.02E-10	***EYA4***[a]	0
14	rs1566129	T	0.41	1.000	0.0037	0.001	2.50E-09	2.53E-09	***NID2****, RTRAF*	0
3	rs3915060	C	0.27	0.983	0.004	0.001	9.70E-09	9.70E-09	***ILDR1***[a]*, CD86, SLC15A2*	0
10	rs10901863	C	0.73	0.934	-0.004	0.001	2.60E-08	2.65E-08	***CTBP2***	0
8	rs7823971	C	0.80	0.991	-0.0043	0.001	2.70E-08	2.68E-08	***RP11-1102P16.1***	0

Table 4.2. Results output from BOLT-LMM and GCTA-COJO for *HAid* association analysis.

Results output from BOLT-LMM and GCTA-COJO. Chr., chromosome; SNP, single nucleotide polymorphism; EA, effect allele; EAF, frequency of effect allele in BOLT-LMM; INFO, quality metric, combination of imputation score and dosage confidence; β, effect size from BOLT-LMM approximation to infinitesimal mixed model; SE, standard error of the effect size; p-value, infinitesimal mixed-effects model association test p-value; pJ-value, p-value from a joint analysis of all the selected SNPs; Genes within 100kb, protein-coding genes that are within 100kb of the lead SNP. The gene in closest proximity to the SNP is emboldened; Dist. to nearest gene (bp), distance in base pairs between SNP and nearest gene, a value of 0 indicates the SNP lies within the gene. a denotes genes previously linked to hearing phenotypes in mice or humans; Two SNPs reached genome-wide significance that are in close proximity to HLA-DQA1 on Chr 6 (Figure 4.4) but were not present in conditional analysis results.

4.2.1.5 Heritability analysis of *HDiff* and *HAid*

SNP heritability estimates (h^2g) were calculated for *HDiff* and *HAid* with BOLT-LMM, as described in - Methods 2.2.3. The h^2g was calculated as 0.117±0.001 *HDiff* for and h^2g and 0.029±0.001 for *HAid*. BOLT-LMM is a tool for linear mixed models' analysis and so this calculation assesses the variance under the assumption that the phenotype is a continuous trait. *HDiff* and *HAid* are qualitative phenotypes and so the estimates require a recalculation to the liability scale. The heritability estimates recalculated to the liability scale are 0.19 and 0.13 for *HDiff* and *HAid* respectively.

4.2.2 Testing the replication of UKBB genetic associations in other samples

An established method to validate GWAS associations is to replicate the association analysis in an independent study sample[282]. Here, replication was sought for the 41 lead SNPs in *HDiff* and the 7 lead SNPs in *HAid* by meta-analysing three independent samples; (i) the remaining northern-European population group in the UKBB cohort (white, non-British Europeans), (ii) TwinsUK, and (iii) the English Longitudinal Study of Ageing (ELSA). Individually, each of these samples are relatively small in size and so a meta-analysis approach was taken in order to increase the statistical power of the replication analysis.

4.2.2.1 Samples and phenotypes used in replication analysis

TwinsUK and the English Longitudinal Study of Ageing (ELSA) are two cohorts that were selected as replication samples. The cohorts were deemed suitable as the majority of the sample subjects are of northern European ancestry and both studies have collected hearing-related questionnaire data that is comparable to that collected by the UKBB. Further descriptions of the two samples are in - Methods 2.1.3 (ELSA) and - Methods 2.1.2 (TwinsUK). The third sample comprised of subjects from the UKBB that are of northern European-ancestry and that were not used in the discovery association analysis, as detailed in - Methods 2.1.1.2. Qualitative phenotypes were derived from questionnaire responses for both the ELSA and TwinsUK samples, as detailed in - Methods 2.1.2.2 and summarised below.

Qualitative phenotypes were derived from questionnaire responses for both the ELSA and TwinsUK samples, as detailed in - Methods 2.1.2.2. In the ELSA sample, using data from wave 7 only, participants that responded "Yes" to both of the following questions; "Do you ever have any difficulties with your hearing?" and "Do you find it difficult to

follow a conversation if there is background noise, such as TV, radio or children playing (using a hearing aid as usual)?" were classified as *HDiff* cases, and participants that responded "No" to both, were classified as *HDiff* controls. As in the UKBB discovery analysis, controls who reported hearing aid use or age <50 were removed, as were any cochlear implant users in the case or control samples. Participants that responded, "Yes most of the time", or "Yes some of the time" to "Whether ever wears a hearing aid" were classified as *HAid* cases and those that responded "No" were classified as *HAid* controls.

In the TwinsUK sample, participants that responded either "Yes, diagnosed by doctor or health professional" or "Yes, not diagnosed by health professional" to the question "Do you suffer from hearing loss?" were classified as *HDiff* cases while those that responded "No" to any of the questions were classified as *HDiff* controls. Participants that responded or indicated "Yes" to either of "Do you wear a hearing aid?" and 'Wearing a hearing aid' were classified as *HAid* cases and those that responded no were classified as *HAid* controls. Twins aged <40 were removed from case and control groups prior to analysis so that the lower age limit was comparable to the UKBB cohort. In order to retain the size of the TwinsUK sample and thus optimise power, controls aged <50 were not removed (as in the discovery *HDiff* UKBB analysis).

The same phenotype classification that was used for the discovery association analysis (described in Chapter 3) was used for the UKBB white non-British European sample, see - Methods 2.1.1.2. (here used for the replication analysis), resulting in *HDiff* n=23,582 and *HAid* n=27,087. Resulting Twins UK samples sizes for association analysis of these traits were *HDiff* n = 3636 and *HAid* n = 3435 and for ELSA, samples sizes for association analysis of these traits were *HDiff* n = 3545 and *HAid* n = 4482. As noted previously, these samples are relatively small by comparison to the UKBB discovery sample and so a meta-analysis approach was taken to increase statistical power to detect an association with the selected SNPs. The meta-analysis sample sizes totalled *HDiff* n = 30,765 and *HAid* n = 35,004.

4.2.2.2 Association analysis in replication samples

Association analysis was conducted on each of the three samples individually, prior to the meta-analysis. Due to intrinsic differences between the samples and datasets, different models were used to test for association in each of the three samples. Genotype and imputation details for TwinsUK and ELSA are listed in are listed in - Methods 2.1.2.3 and 2.1.3.3 respectively. A meta-analysis of all three association analyses was performed on the 48 lead SNPs identified in the two discovery GWAS (*HDiff* and *HAid*). A fixed-effect

inverse-variance weighted meta-analysis was conducted using METAL[253] version 2011-03-25.

Replication was sought for all leading SNPs identified by conditional and joint analysis (41 for *HDiff* and 7 for *HAid*), and so Bonferroni corrections were applied in order to account for multiple testing. This resulted in the Bonferroni-corrected thresholds of significance being 0.05/41=0.0012 for *HDiff* and 0.05/7=0.00714 for *HAid*. Two intronic SNPs, rs759016271 in *ZNF318* and rs1566129 in *NID2*, reached significance in the *HDiff* replication analysis (P<0.0012), and a further intronic SNP, rs4597943 in *ARHGEF28* replicated in *HAid* analysis at the significance threshold (P<0.00714). An additional 14 SNPs reached nominal significance (Table 4.3and Table 4.4).

Replication association analysis was conducted on the lead SNP at each locus. This was to avoid obtaining an over-conservative significance threshold; the greater the number of SNPs tested, the more conservative the Bonferroni level of significance. In both replication samples, all lead SNPs were present, i.e. genotyped, or imputed. It was therefore deduced that if the lead SNP was not the 'causal' SNP at the locus, the lead SNP would feature an inflated test statistic due to it being in LD with the causal SNP. This is because the two replication samples are of the same ancestral population subgroup as the discovery analysis and would therefore exhibit largely similar LD structures. This would remain the case if the 'causal' SNP did not feature on the genotyping platform or imputation panel in any of the samples.

4.2.2.3 Estimating the statistical power of the replication analysis

Power calculations were performed to assess the power of this replication sample to replicate the initial findings from the discovery association analysis[282]. Individual power calculations were performed for each of the 48 lead SNPs (for equation, see - Methods 2.2.4.2). The significance value was Bonferroni corrected (0.05/41 and 0.05/7) for both traits and thus was set at p<0.0012 for *HDiff* and p<0.00714 for *HAid*.

Based on these calculations, the meta-analysis sample is estimated to have >80% power to replicate 6 of the 41 *HDiff* lead SNPs at a Bonferroni-corrected significance level, p<0.0012. For 39 of the 41 SNPs, the sample was estimated to provide >80% to replicate the associations at a nominal level of significance, p<0.05. As for the replication of *HAid* lead SNPs, this sample was estimated to have <80% power to replicate all 8 lead SNPs at either Bonferroni-corrected or nominal thresholds of significance. The results of these power calculations are displayed alongside the replication summary statistics below, in Table 4.3 and Table 4.4.

SNP	A1	A2	Weight	Zscore	P-value	Dir.	Rep. pwr. p<0.05	Rep. pwr. p<0.0012
rs759016271	a	Agtagtcc Actttttctt ctttgcctg	29866	4.556	5.20E-06	+++	0.900	0.504
rs1566129	t	c	30894	3.534	0.00041	+++	0.661	0.195
rs36062310	a	g	30868	2.974	0.002938	+--	0.928	0.575
rs143282422	a	g	30274	2.652	0.007996	+-+	0.492	0.098
rs12225399	c	g	29802	2.624	0.008688	+++	0.610	0.160
rs6597883	t	c	30274	2.585	0.009748	++-	0.564	0.133
rs62033400	a	g	30851	2.513	0.01198	+++	0.600	0.153
rs7951935	t	g	29802	2.48	0.01312	++-	0.820	0.360
rs141403654	a	t	29802	-2.392	0.01674	---	0.477	0.091
rs35186928	a	g	29866	2.348	0.01885	+-+	0.785	0.314
rs17671352	t	c	30852	2.228	0.02587	++-	0.520	0.110
rs217289	a	g	29866	2.227	0.02597	+-+	0.524	0.112
rs62188635	t	c	30377	-2.189	0.02859	---	0.583	0.144
rs76837345	a	g	30318	-2.158	0.0309	---	0.499	0.101
rs55635402	a	g	29802	2.132	0.03301	+++	0.585	0.145
rs10824108	t	g	30274	1.982	0.04752	+-+	0.541	0.120
rs2236401	t	c	29866	1.873	0.06102	++-	0.561	0.131
rs5756795	t	c	30868	-1.696	0.08981	--+	0.679	0.209
rs7525101	t	c	30389	1.636	0.1018	+++	0.506	0.104
rs10475169	a	c	30356	-1.603	0.1088	-+-	0.516	0.108
rs4948502	t	c	30274	1.574	0.1155	+-+	0.553	0.127
rs9691831	a	g	30717	-1.512	0.1305	-+-	0.492	0.098
rs12938775	a	g	30852	-1.481	0.1385	-+-	0.509	0.105
rs4611552	t	c	30753	-1.434	0.1515	---	0.489	0.096
rs13093972	a	g	30098	-1.413	0.1577	--+	0.525	0.113
rs6453022	a	c	30356	1.327	0.1846	+--	0.912	0.532
rs835267	a	g	30274	1.17	0.242	++-	0.555	0.128
rs35414371	a	t	30856	1.163	0.245	+-+	0.657	0.192
rs9493627	a	g	29866	1.018	0.3086	+++	0.724	0.248
rs12027345	a	g	30389	-0.986	0.3244	--+	0.537	0.119
rs3890736	a	g	30318	-0.706	0.4805	+--	0.495	0.099
rs4947828	t	g	30717	-0.562	0.574	+--	0.571	0.137
rs13277721	a	g	30318	0.527	0.5983	+--	0.590	0.147
rs12552	a	g	30676	0.489	0.6249	+-+	0.480	0.093
rs6890164	a	g	30356	0.298	0.7658	+--	0.877	0.455
3:182069497_TA_T	t	ta	30098	0.298	0.766	++-	0.629	0.172
rs132929	a	g	30868	0.234	0.815	++-	0.731	0.256
rs10927035	t	c	30389	0.219	0.8268	+-+	0.476	0.091
rs34442808	t	ta	30356	0.218	0.8277	-++	0.560	0.131

SNP	A1	A2	Weight	Zscore	P-value	Dir.	Rep. pwr. p<0.05	Rep. pwr. p<0.0012
rs9366417	a	g	29866	0.137	0.8911	+-+	0.490	0.097
rs62015206	t	c	30459	-0.071	0.9437	--+	0.522	0.111

Table 4.3. Summary statistics for HDiff phenotype from the replication meta-analysis of the white non-British UKBB sample, TwinsUK and ELSA.

SNP, SNP ID; A1, the first allele for this marker in the first file where it occurs; A2, the second allele for this marker in the first file where it occurs; Weight, the sum of the individual study weights (N) for this marker; Z-score, the combined z-statistic for the marker; P-value, meta-analysis p-value; Direction, direction of effect for each study ordered: white non-British UKBB sample, TwinsUK, ELSA; Replication power at p<0.05, estimated power to identify a replicated association at nominal significance; Rep. pwr. p<0.0012, estimated power to detect an association at significance threshold 0.05/41 = 0.0021, p<0.0012.

SNP	A1	A2	Weight	Zscore	P-value	Dir.	Rep. pwr. p<0.05	Rep. pwr. p<0.00714
rs4597943	t	g	34475	2.833	0.004608	+++	0.695	0.413
rs7823971	a	c	34919	-1.886	0.05934	---	0.541	0.265
rs3915060	t	c	34359	-1.664	0.09605	--+	0.560	0.281
rs1566129	t	c	35139	0.784	0.4333	++-	0.602	0.318
rs10901863	t	c	32251	0.638	0.5234	+--	0.510	0.240
rs9321402	a	g	35101	0.58	0.5622	++-	0.649	0.364
rs9677089	a	c	34727	-0.537	0.5915	--+	0.653	0.368

Table 4.4. Summary statistics for HDiff phenotype from the replication meta-analysis of the white non-British UKBB sample, TwinsUK and ELSA.
SNP, SNP ID; A1, the first allele for this marker in the first file where it occurs; A2, the second allele for this marker in the first file where it occurs; Weight, the sum of the individual study weights (N) for this marker; Z-score, the combined z-statistic for the marker; P-value, meta-analysis p-value; Direction, direction of effect for each study ordered: white non-British UKBB sample, TwinsUK, ELSA; Replication power at p<0.05, estimated power to identify a replicated association at nominal significance; Rep. pwr. p<0.0012, estimated power to detect an association at significance threshold 0.05/7 = 0.00714, p<0.00714.

4.2.2.4 Replication of previously published ARHI GWAS findings

In addition to the meta-analysis replication, it was investigated whether any associations that are highlighted by authors in previous ARHI GWAS, are replicated here in the *HDiff* or *HAid* discovery White British sample. Previous ARHI GWAS have identified significant associations between 5 gene loci (*SIK3*[111], *PCDH20*[112], *SLC8A3*[112], *ISG20*[113] and *TRIOBP*[113]) and ARHI phenotypes, while all other studies reported the most highly associated SNPs.

Only two of the previously highlighted SNPs, located in close proximity to *ISG20* and within *TRIOBP*, were significant in the *HDiff* analysis at a Bonferroni-corrected level of significance (p<7.50E-07). These two associations were originally identified in a GWAS that used phenotypes that were derived from electronic health records (Hoffman *et al.*, 2016[113]). No other lead variants from previous ARHI genetic studies were replicated at nominal level in this analysis, including the first reported ARHI associated gene variant in *GRM7*[109]. The summary statistics from *HDiff* and *HAid* for the SNPs highlighted in previous analysis are listed in Table 4.5. The legend for Table 4.5 is presented on the following page, is:

Table 4.5. Summary statistics from *HDiff* and *HAid* GWAS analysis, at SNPs highlighted in previous adult hearing loss GWAS.

Study, publication of previous finding; Gene, gene highlighted in the referenced publication as the lead SNP is either located in the gene region or in close proximity; SNP, single nucleotide polymorphism; CHR, Chromosome; BP, base position; A1, effect allele in analysis; A0, reference allele; INFO, quality metric, combination of imputation score and dosage confidence; UKBB pheno, phenotype used in this study; A1FREQ, frequency of effect allele in analysis sample; BETA, effect size from BOLT-LMM approximation to infinitesimal mixed model; SE, standard error of the effect size; p-value, infinitesimal mixed model association test p-value. *The SNP rs58389158 was not included in this analysis, but the SNP rs5756795, which is in complete LD with this SNP in the British population is listed here and referenced in the previous study.

Variant highlighted in previous study				Summary statistics from *HDiff* and *HAid* analysis in the UKBB cohort								
Citation	Gene	SNP	CHR	BP	A1	A0	INFO	UKBB pheno	A1FREQ	BETA	SE	p-value
Friedman *et al.*, 2009[109]	*GRM7*	rs11928865	3	7155702	T	A	0.989	*HDiff*	0.741	0.0016	0.0015	0.28
								HAid	0.742	-0.0014	0.0007	0.05
Van Laer *et al.*, 2010[107]	*IQGAP2*	rs457717	5	75920972	A	G	0.986	*HDiff*	0.326	0.0013	0.0014	0.34
								HAid	0.325	-0.0006	0.0007	0.37
	GRM7	rs161927	3	7838242	G	A	0.988	*HDiff*	0.134	0.0038	0.0019	0.05
								HAid	0.136	-0.0002	0.0009	0.86
Girotto *et al.*, 2011[114]	*DCLK1*	rs248626	5	141097725	A	G	1.000	*HDiff*	0.251	0.0018	0.0015	0.23
								HAid	0.252	-0.0003	0.0007	0.71
	KCNMB2	rs4603971	3	177902467	G	A	0.992	*HDiff*	0.934	-0.0015	0.0027	0.58
								HAid	0.934	0.0006	0.0012	0.63
	CMIP	rs898967	16	81566780	C	T	0.981	*HDiff*	0.476	0.0010	0.0013	0.45
								HAid	0.476	0.0002	0.0006	0.76
	GRM8	rs2687481	7	125869122	G	T	0.998	*HDiff*	0.811	-0.0018	0.0017	0.28
								HAid	0.810	0.0012	0.0008	0.14
Nolan *et al.*, 2013[92]	*ESSRG*	rs2818964	1	216682448	G	A	0.978	*HDiff*	0.366	-0.0015	0.0014	0.27
								HAid	0.366	0.0004	0.0006	0.55
Wolber *et al.*, 2014[111]	*SIK3*	rs681524	11	116748314	T	C	0.992	*HDiff*	0.927	-0.0010	0.0026	0.71
								HAid	0.928	0.0018	0.0012	0.13
Vuckovic *et al.*, 2015[112]	*PCDH20*	rs78043697	13	62467039	T	C	0.995	*HDiff*	0.928	0.0000	0.0025	1.00
								HAid	0.928	0.0010	0.0012	0.38
	SLC28A3	rs7032430	9	86714002	C	A	0.959	*HDiff*	0.782	-0.0013	0.0016	0.43
								HAid	0.783	-0.0001	0.0008	0.91
Fransen *et al.*, 2015[38]	*ACVR1B*	rs2252518	12	52381026	C	A	0.996	*HDiff*	0.739	-0.0010	0.0015	0.50
								HAid	0.739	0.0001	0.0007	0.85
	CCBE1	rs34175168	18	57180682	G	A	0.990	*HDiff*	0.986	0.0112	0.0056	0.04
								HAid	0.986	-0.0009	0.0026	0.74
Hoffman *et al.*, 2016[113]	*ISG20*	rs4932196	15	89253268	T	C	1.000	*HDiff*	0.809	0.0085	0.0017	4.60E-07
								HAid	0.809	0.0039	0.0008	6.40E-07
	TRIOBP	rs5756795*	22	38122122	T	C	1	*HDiff*	0.539	-0.0092	0.0013	5.10E-12
								HAid	0.538	-0.0027	0.0006	1.60E-05

4.2.3 *In silico* functional interpretation of genetic association analysis results

The final sections of this chapter consist of *in silico* functional annotation of the results from the association analysis in section 4.2.1. In each of the following sections, methods are used to translate the SNP associations that were identified and validated in the previous sections, into putative findings relating to their biological relevance and multi-trait effects.

4.2.3.1 Gene-set analysis

Gene-set analysis is a method that is commonly used for translating GWAS SNP associations into biologically relevant knowledge. Associated SNPs can be mapped to genes based on SNP location and function. These genes are then compared to pre-defined gene sets. Gene sets that show significant enrichment of the mapped genes, can thus be putatively linked to the trait of interest. Gene sets can allude to biological mechanisms as gene sets are defined by common functions and, or expression patterns.

This is a particularly useful method in the analysis of polygenic traits because the combined effects of the associated variants (with individual small effect sizes) can be studied. In addition, by studying the combined effects, the threshold for significance is also lowered. The analysis was performed here using FUMA[233], in two distinct steps; 'SNP2GENE' and 'GENE2FUNC', where the results in the first step are used to calculate the latter. Firstly, gene association p-values are quantified based on associations of SNPs within 10kb of gene transcripts. Secondly, this group of genes are tested for enrichment in pre-defined gene sets, see - Methods 2.3.1.

The gene sets and gene ontology (GO) terms that are used to classify the gene sets are from MSigDB v5.2 (http://software.broadinstitute.org/gsea/index.jsp). GO annotations include three groups; molecular function, cellular component and biological process. GENE2FUNC analysis was performed for both *HDiff* and *HAid.* In the SNP2GENE function, SNPs from the association analysis were mapped to 113 protein coding genes that were within 10kb of SNP location. For *HDiff*, significant gene sets are displayed in Figure 4.6; 28 GO biological functions were significantly enriched with genes from the 113 prioritised genes from *HDiff* analysis. As the number of genes within gene sets differ, and the number of genes that are enriched in each set are differ. The pink bar chart in Figure 4.6 displays the proportion of overlapping genes per gene set, while the blue bar chart visualises the p-value for enrichment in each gene set. The yellow blocks on the graph demonstrate which genes (x-axis) are included in each gene set (y-axis).

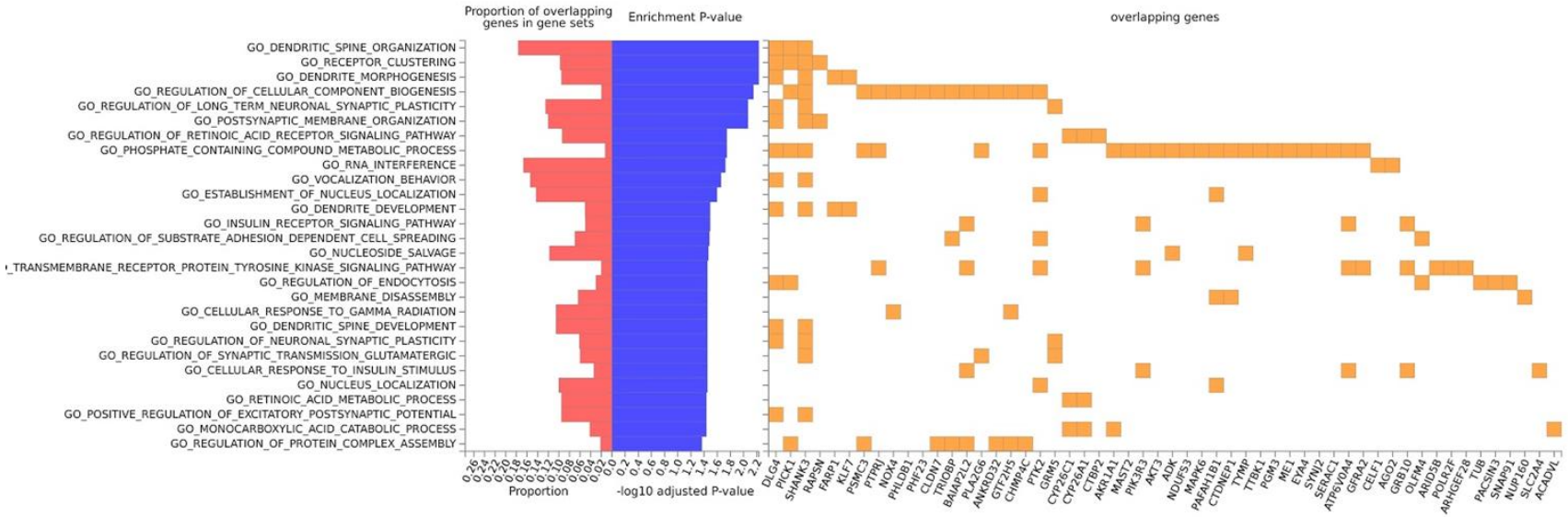

Figure 4.6. Bar chart of gene-set analysis with *HDiff*.

GO Biological processes (bottom), displaying the proportion of overlapping genes within assigned gene sets and the p-value for enrichment for each function. GO biological functions with p-values <0.05 are displayed on this graph.

The three GO biological functions with the most significant p-values are Dendritic spine organisation, Receptor clustering and Dendrite morphogenesis, all with an adjusted enrichment p-value of p=5.87e-3. All three of these processes are concerned with relaying or receiving information via either neuron or receptor structure and function. Based on the known structure of the auditory system, the function of associated genes can therefore be hypothesised. The biological processes that individual genes are associated with, and the interactions between genes within the same groups can be investigated further.

4.2.3.2 Genetic correlation analysis between *HDiff* and >700 traits

Genetic correlation analysis is a method that can be used to investigate relationships between traits, based on the proportion of their variance that is due to shared genetic variants. Genetic correlations calculated via this method range from -1 rg to 1 rg, where 0 indicates that the genetic causes of trait variance are entirely independent and 1 indicates a complete correlation while a negative score indicates a negative correlation. Using publicly available summary statistics from several hundred GWAS analyses, and the summary statistics from *HDiff* association analysis, it has been possible to estimate the genetic correlation between *HDiff* and 764 different traits (publicly available GWAS summary statistics on LD hub[251,255] see - Methods 2.3.2.).

Following a Bonferroni correction (0.05/764, p< 6.54E-05), 153 traits display a significant correlation with *HDiff*. The trait with the highest significant correlation to *HDiff* is 'Tinnitus: Yes_ now most or all of the time'. This trait represents a sample that report experiencing tinnitus symptoms 'now or most of the time' that can be described as 'noise (such as ringing or buzzing) in your head or in one or both of your ears that lasts for more than five minutes at a time?' The data used in the correlation analysis with *HDiff*, consists of results from a GWAS run at the Neale lab that used UKBB tinnitus questionnaire responses as a phenotype (http://www.nealelab.is/)[283].

41 of the significant traits have an rg< -0.3 or rg> 0.3 and are listed in Table 4.6. These traits are grouped into five categories based on similarities between phenotype symptoms: breathing difficulties, health report/subject wellbeing, hearing, low mood/depression and pain. Dental pain (toothache and painful gums), frequency of tiredness/lethargy and other eye problems also had a significant rg>0.4 with *HDiff.* The majority of these traits have symptoms of depression and low mood, while three are related to breathing difficulties and seven relate to reporting the experience of different pain types.

Trait	rg	se	p	Group
Wheeze or whistling in the chest in last year	0.3137	0.0306	1.35E-24	Breathing difficulties
Shortness of breath walking on level ground	0.3228	0.0429	5.60E-14	Breathing difficulties
Bring up phlegm/sputum/mucus on most days	0.342	0.0679	4.74E-07	Breathing difficulties
Long-standing illness_ disability or infirmity	0.3763	0.0305	5.48E-35	Health report/Subjective wellbeing
Overall health rating	0.3156	0.0266	2.09E-32	Health report/Subjective wellbeing
Health satisfaction	0.3405	0.0374	8.47E-20	Health report/Subjective wellbeing
Other serious medical condition/disability diagnosed by doctor	0.3203	0.0409	4.80E-15	Health report/Subjective wellbeing
Subjective well being	-0.3257	0.0421	1.06E-14	Health report/Subjective wellbeing
Other eye problems	0.4311	0.0687	3.59E-10	Health report/Subjective wellbeing
Had major operations	0.3196	0.06	9.84E-08	Health report/Subjective wellbeing
Illnesses of siblings: None of the above (group 2)	-0.3436	0.0702	9.94E-07	Health report/Subjective wellbeing
Former alcohol drinker	0.3136	0.0701	7.56E-06	Health report/Subjective wellbeing
Tinnitus: Yes_ now most or all of the time	0.6	0.0562	1.40E-26	Hearing
Loud music exposure frequency	0.3224	0.0583	3.20E-08	Hearing
Frequency of tiredness / lethargy in last 2 weeks	0.4089	0.029	2.79E-45	Low mood /depression
Neuroticism score	0.315	0.0257	1.94E-34	Low mood /depression
Miserableness	0.3283	0.0273	2.69E-33	Low mood /depression
Seen doctor (GP) for nerves_ anxiety_ tension or depression	0.3447	0.0298	5.90E-31	Low mood /depression
Frequency of depressed mood in last 2 weeks	0.3361	0.0318	4.30E-26	Low mood /depression
Guilty feelings	0.3258	0.0309	4.98E-26	Low mood /depression
Loneliness_ isolation	0.3233	0.0333	2.99E-22	Low mood /depression
Frequency of unenthusiasm / disinterest in last 2 weeks	0.3239	0.0348	1.35E-20	Low mood /depression
Frequency of tenseness / restlessness in last 2 weeks	0.3046	0.0336	1.26E-19	Low mood /depression
Ever depressed for a whole week	0.371	0.041	1.45E-19	Low mood /depression
Illness_ injury_ bereavement_ stress in last 2 years: Financial difficulties	0.3098	0.0368	3.70E-17	Low mood /depression

Trait	rg	se	p	Group
Depressive symptoms	0.3314	0.0409	5.69E-16	Low mood /depression
Ever unenthusiastic/disinterested for a whole week	0.3445	0.0432	1.53E-15	Low mood /depression
Happiness	0.3148	0.0408	1.27E-14	Low mood /depression
Financial situation satisfaction	0.3437	0.0473	3.87E-13	Low mood /depression
Family relationship satisfaction	0.3118	0.0433	6.12E-13	Low mood /depression
Ever highly irritable/argumentative for 2 days	0.3192	0.0457	2.78E-12	Low mood /depression
Insomnia	0.3211	0.0484	3.31E-11	Low mood /depression
Illness_ injury_ bereavement_ stress in last 2 years: Serious illness_ injury or assault to yourself	0.3161	0.0485	7.36E-11	Low mood /depression
Neuroticism	0.3125	0.0724	1.60E-05	Low mood /depression
Chest pain or discomfort	0.3823	0.0341	3.60E-29	Pain
Pain type(s) experienced in last month: None of the above	-0.3225	0.0295	9.62E-28	Pain
Pain type(s) experienced in last month: Neck or shoulder pain	0.3686	0.0367	9.34E-24	Pain
Pain type(s) experienced in last month: Stomach or abdominal pain	0.3409	0.0418	3.73E-16	Pain
Pain type(s) experienced in last month: Hip pain	0.3079	0.0394	5.25E-15	Pain
Mouth/teeth dental problems: Toothache	0.428	0.0699	9.02E-10	Pain
Mouth/teeth dental problems: Painful gums	0.4514	0.0749	1.71E-09	Pain

Table 4.6. Genetic correlation results with *HDiff* phenotype.

This table lists traits that had significant correlations with the *HDiff* phenotype and an rg < -0.3 or rg > 0.3. The traits are ordered by group, and by p-value order within each group. Trait; trait as listed on LD hub, rg; genetic correlation, se; standard error of rg, p-value; p-value corresponding to rg, Group; traits were grouped based on common or similar symptoms and groups named according to common symptoms.

4.3 Discussion

Phenotypes and association analysis

The *HDiff* and *HAid* association analyses are the two largest hearing related GWAS to date. Over 2,320 genome-wide significant SNPs were identified, representing 44 independent associations with self-reported adult hearing difficulty in participants aged 40-69 years. The findings represent a nine-fold increase in independent, genome-wide significant genomic loci for ARHI-related phenotypes. A quarter of the loci identified have already been implicated in hearing conditions, while the remaining three quarters are novel associations with auditory function.

The two studies demonstrate the utility of using self-report measures to successfully identify significant associations with hearing phenotypes in genetic population analysis. The studies demonstrate the scale that is required in order to achieve sufficient power to identify multiple variants of small effect sizes that cause complex traits. This work also provides the field with surrogate measures to assess common HL, that are practical to obtain on a large scale. There are some potential limitations to using questionnaires to assess hearing ability, as detailed in Chapter 3, meaning that the analysis may therefore have reduced power due to a bias in false-positive cases and, or false-negative controls.

To reduce the likelihood of associations being driven by highly penetrant variants that cause forms of hearing impairment that are distinct to ARHI, rare variants were excluded from the analysis. A minor allele frequency threshold of >0.01 was implemented to exclude risk variants for forms of congenital hearing loss, as this is lower than the prevalence of congenital hearing loss in the UK population. Secondly, participants who selected 'I am completely deaf' in the UKBB questionnaire were excluded. As discussed in chapter 3, the age distribution of this subset of individuals that responded, 'I am completely deaf', was not significantly different from age distribution of the entire sample and so likely does not predominantly represent ARHI cases but rather individuals with other forms of hearing loss.

Four loci were associated with both *HDiff* and *HAid* at genome-wide significance. Three further loci were associated with *HAid* yet not *HDiff*, and 37 were associated with *HDiff* yet not with *HAid*. The disparity in the number of findings is likely to be due in part to the reduced statistical power of *HAid* in comparison to *HDiff*. The case prevalence was 5% in *HAid* and 35% *HDiff,* and the overall sample sizes were n=253,918 and n=250,389 for *HAid* and *HDiff* respectively. This meant that the statistical power of *HAid* was much reduced compared to *HDiff*, due to the case: control ratio, a greater case: control ratios yield greater statistical power in association analyses. SAIGE[228] provides an alternative to the BOLT-

LMM[226] model that can account for case: control imbalance in association analysis. However, BOLT-LMM was selected as it was the only available method that could compute a relationship matrix for a sample of this size. The matrix is necessary in order to account for the population structure within the UKBB, as discussed in 'The UK Biobank cohort' subsection of 1.4,.

In addition to the disparity between the number of significant associations, there were multiple loci that were distinctly associated with either *HAid* or *HDiff*. This may indicate that the two phenotypes depict subtly distinct pathologies. Frequent hearing aid users likely experience a relief in symptoms with the use of the amplification device, indicative of an inner ear pathology. On the other hand, individuals that have trouble when listening to speech when background noise is present, may experience a pathology that affects both the peripheral and central auditory systems. HDiff may therefore depict a hearing impairment of mixed pathologies or one that at least encompasses the central auditory system, akin the hypothesis that most cases of ARHI present as a mixed pathology.

Separately, GWAS are commonly adjusted for covariates where appropriate. The aim of these adjustments is to account for confounding, where variation in the outcome (trait) is due to the effect of the covariate. Confounding factors can either (i) predict the outcome in the absence of an exposure, (ii) be associated with but not act as a surrogate for the outcome or (iii) can interact with both the exposure and the outcome. An adjustment for these factors can increase the power to identify genetic variants that have a direct effect on the outcome (trait)[284]. Conversely, adjustment for covariates can reduce statistical power when the tested genetic variants have an effect on the covariate, or when a correlation between the trait and the covariate is not fully explained by the direct effect of the covariate on the trait.

HDiff and *HAid* were adjusted for the following covariates: age, sex, UKBB genotyping platform and UKBB PCs 1-10. However, numerous social and lifestyle factors are also thought to contribute to hearing deterioration, as discussed in chapter 2), and which are likely to have an effect on the phenotypes used in this study. For example, hearing aid use in this sample has previously been shown to be significantly associated with Townsend score in the UKBB; those with 'poor' hearing and who wear hearing aids have a significantly lower Townsend score (less deprived) than those who do not[123]. Therefore, the statistical power may be increased by adjusting for covariates such as this. However, the relationship between Townsend score and HL or hearing aid use is not fully understood; it is likely that there are multiple factors encompassed in the Townsend score which have different effects on the risk of developing a hearing loss, the perception of symptoms and the prescription of hearing aids and the frequency of use. Here no further covariates were included, though factors that are associated with or hypothesised to interact with the *HDiff* and *HAid* ought

to be taken into consideration if results are used to estimate population prevalence or to inform future implementation of treatments and diagnostics.

Heritability analysis of *HDiff* and *HAid* in the UKBB

The SNP-based heritability estimates for *HDiff* (h^2_g =19%) and *HAid* (h^2_g =13%), are at the mid-lower end of previous heritability estimates for ARHI[100,103,285], see Table 1.3. This is not unexpected as h^2_g estimates are lower if not equal to h^2 and H^2, as h^2_g is an estimate only of the additive effects of the genotyped SNPs that are included in the analysis[286]. Secondly, phenotypes used to calculate these previous estimates are also predominantly derived from PTA scores, and as discussed, will likely have a different proportion of variance attributable to genetic causes, than the phenotypes used here.

A previous estimate that was calculated with self-report measures however, was also higher than these two estimates, at 40% (95% CI=19-52%). The sample that was used to calculate this estimate was made of up of a subset of a Danish Twin Registry[119], and the narrow-sense heritability estimate in theory assess additive effects of the entire genome, rather than just the genotyped SNPs in the model used here to calculate h^2_g. Furthermore, narrow sense heritability may be susceptible to inflation due to genotype and environmental factors within families that provide a commonality in addition to additive genetic effects[287].

The most recent study to calculate SNP heritability for an ARHI-related trait was the GWAS by Hoffman, 2016, which reported the estimated heritability for all genome-wide SNPs to be 8.7% (95% CI 2.9%-14.4%). Again this method differs to that used to calculate the estimates for *HDiff* and *HAid*, as the Hoffman calculation included additive effects only from SNPs that reached genome-wide significance, while *HDiff* and *HAid* estimates are based on additive effects of all SNPs in the GWAS analysis, regardless of the association p-value. Secondly, the prevalence that was stipulated in order to re-calculate the estimates to the liability scale, differed for each of these three estimates. The Hoffman sample stipulated a prevalence of 12.5% (the prevalence in the non-Hispanic whites in the sample cohort), while the prevalence for *HDiff* and *HAid* are 35% and 5% respectively.

Future studies are likely to add context to the SNP heritability estimates presented in this chapter. This may be by including a greater number of variants and, or including rare variants not currently included in association analyses and which may have greater effect sizes. For this analysis, a SNP MAF of 0.01 was implemented and, following imputation and further QC, 9,740,198 SNPs were included in the *HDiff* and *HAid* analyses. Although this is a relatively large proportion of the SNPs in the genome in comparison to previous ARHI studies, a less conservative MAF and more dense genotyping and imputation will allow for a much larger number of SNPs to be tested for association. In time, it may also be possible

to accurately incorporate epigenetic and epistatic effects and to reduce the effects of confounding factors. As noted in the Introduction p.60, GWAS findings generally only explain 20-30% of the trait variance. Further work is therefore likely to result in an increased, and more accurate, heritability estimate than that presented here, which only compasses the effects of SNPs included in these GWAS.

Replication of genetic association analysis of *HDiff* and *HAid*

Associations at three of the lead SNPs were successfully replicated at a Bonferroni-corrected level of significance, while an additional 15 SNPs reached nominal significance. Despite being a meta-analysis of three independent samples, the replication sample, was a magnitude smaller in size than the discovery sample. The replication sample therefore had less statistical power than the discovery sample, as is demonstrated by the power calculations. The calculations predict the sample to have >80% power to replicate 6 SNPs at a Bonferroni-corrected level of significance, and 39 at nominal significance.

The observed replication rate however, was half that of the predicted rate (based on these calculations). Assuming that the disparity is not due to false positive findings in the discovery sample, factors other than a reduction in statistical power may be causing the difference between predicted and observed replication rates. Two possible factors are the subtle differences in the questions used to derive the phenotypes in each of the three samples, along with the different models used to test for association may have reduced the power of analysis. The lack of replication currently observed across hearing genetic association studies does not invalidate the initial associations, but rather supports alternative methods of validation via functional analysis, such as by demonstrating a mode of action in auditory mechanisms or, even better, pathogenic consequences when the gene is dysregulated.

Replication of previous ARHI GWAS findings

In addition to the replication meta-analysis, it was investigated whether any SNPs that were highlighted in previous ARHI, are replicated in the *HDiff* or *HAid* sample. Two of the fifteen SNPs that were highlighted in previous studies are replicated in the *HDiff* sample at a Bonferroni-corrected level of significance; a variant that is in close proximity to *ISG20*, and a variant in the *TRIOBP* transcript. This is evidence to support these two previous findings, but does not however invalidate the remaining, previous significant findings. The 15 SNPs were selected from the 8 previous ARHI GWAS, 7 of which used phenotypes derived from PTA thresholds; only Hoffman (the study that identified the *ISG20* and *TRIOBP* associations) used self-report measures to derive the phenotypes, see Table 1.5.

Disparities in findings were also present when different phenotype measures were used to test for association, even within the same population sample. Both Hoffman and Girotto used the Genetic Epidemiology Research on Adult Health and Ageing (GERA) cohort yet of the SNPs highlighted in Girotto[114], none were replicated in the Hoffman analysis[113]. One used a PTA-derived phenotype while the other used electronic health records and self-report measures. Inconsistent phenotyping and the lack of large cohorts (data available for high-powered analysis) is currently a general limitation of hearing research, both for the identification and replication (and thus validation) of genetic risk variants. As a result, validation methods rely heavily on functional analysis.

Findings from genetic association analysis that are linked with forms of congenital deafness

There has been speculation as to whether ARHI susceptibility variants lie within genes that are known to cause congenital deafness. If so, such variants would be expected to be less deleterious and expressed post-development, resulting in more mild forms of later-onset hearing impairment than as seen in forms of congenital deafness. The alternative hypothesis is that ARHI susceptibility genes are entirely novel to hearing and affect mechanisms distinct to those involved in congenital HL. These results indicate that both hypotheses are partially correct; a quarter of the lead SNPs in *HDiff* and *HAid* reside in or are in close proximity to known hearing genes. Only one of the lead SNPs is within a gene transcript that has previously been associated with ARHI (rs5756795, in *TRIOBP*) while nine loci have previously been linked to some form of hearing loss either in mouse models or humans (Table 4.1 and Table 4.2). The remaining three quarters reside in loci that have no previous links to hearing function or pathologies. Within this group of novel associations however, there may be genetic variants that cause other forms of HL but have not yet been identified.

Four significant gene loci are common to both *HDiff* and *HAid* analyses; *EYA4, NID2, ARHGEF28* and *CTBP2*. Variants within *EYA4* have been reported in autosomal dominant non-syndromic hearing loss[204,288,289], while *NID2* and *ARHGEF28* are novel associations with hearing phenotypes and auditory function. In addition to these, *CDH23* and *BAIAP2L2* have been linked to forms of syndromic deafness[290,291] while *EYA4*[204,288,289], *TRIOBP, ILDR1* and *LMX1A* have been linked to non-syndromic HL[16,113,292]. Four further loci contain genes that have been implicated in hearing function, and these association results provide the first link between these genes (*CTBP2*, *SYNJ2*, *SPTBN1* and *TUB*) and a human pathology.

CTBP2 encodes C-terminal Binding Protein 2 a critical protein component of the inner ear hair cell pre-synaptic ribbon[293]. The function of the ribbon synapses is to facilitate the exocytosis function of synaptic vesicles, the organisation of which is highly specialised for

precise temporal neurotransmission along the organ of Corti[294]. *SYNJ2* encodes Synaptojanin 2, an inositol phosphatase that is known to function in recycling neurotransmitter vesicles[295]. Its role in auditory function is yet to be established, though Synj2^{mozart} mice display a progressive HL in conjunction with a deterioration of sensory HCs. *SPTBN1* encodes βII-Spectrin, which has a role in anchoring HC rootlets in the cuticular plate, found at apex of sensory HCs, by providing mechanical support in conjunction with αII-Spectrin[296].

Tub^{tub} mice exhibit progressive hearing loss and retinal degeneration in the absence of further neurological phenotypes, as observed in USH1[297]. Cochlea from Tubby$^{-/-}$ mice have an accelerated outer and inner hair cell loss when compared to Tubby$^{-/+}$ and Tubby$^{+/+}$ mice and exhibit spiral ganglion cell loss[298]. The significance of identifying 10 known hearing loci is discussed in detail in chapter 5, along with further analysis of the associations with *NID2* and *ARHGEF28*.

Gene-set analysis

Gene-set analysis is a valuable tool for interpreting GWAS-derived SNP associations. In common complex traits, each variant that is associated with the trait has a relatively small effect size and often an ambiguous function. So, individually, SNP associations cannot generally reveal pathogenic pathways and mechanisms. By combining the effects and targets of associated SNPs, trait-relevant pathways can be identified, and pathogenic hypotheses can be derived, as discussed in section 1.4. Pathway analysis presented here resulted in significant findings with the *HDiff* phenotype and the results of GO Biological components are presented. A possible reason for the disparity in the number of significant GO processes, functions and components may be due to the quantity of significant associations in the original analysis. As the *HDiff* GWAS had greater power to identify significant associations (higher case: control ratio than *HAid*), there were a greater number of SNPs and thus genes included in MAGMA analysis; the MAGMA *HDiff* input used 113 genes while *HAid* input used only 67 genes based on the same inclusion threshold parameters. Therefore, when using the *HAid* data, fewer gene sets will have been included in the permutations and thus could not reach significance.

The most significant GO biological process was dendritic spine organisation, while three additional significantly enriched relate to dendritic morphogenesis and development. This is in line with early theories of ARHI pathogenesis (sensory and neuronal) and where dendritic (and synaptic) density and morphology has been shown to change with age in the inferior colliculus (principal midbrain nucleus of the auditory pathway)[299]. The second most significant GO biological process was receptor clustering, one of a number of the significant GO biological terms that relate to properties of receptors or synapses. There are a number

of synaptic processes throughout the auditory pathway, from HCs to higher auditory structures. Furthermore, a recent study has shown that synapse degeneration is observed in the auditory cortex and hippocampus CA3 region in conjunction with hearing threshold increase and onset of profound hearing loss in C57BL/6J mice[300]. The results of gene-set enrichment analysis can vary widely based on the method selected, model parameters used and the selected reference dataset[301]. As the knowledge of gene sets is rapidly increasing, the composition of gene sets is continually evolving so these results are artificially limited to data that has been used to curate the gene sets.

Significant genetic correlations between *HDiff* and multiple traits

As discussed in Chapter 1 'Associated conditions', epidemiological studies commonly report associations between different traits in population samples. However, these reports can be subject to reverse causation and a higher number of correlated cofounders[184]. Genetic correlation analysis is a method that can be used to identify shared genetic variance between traits, and that is less susceptible to reverse causation and the effects of correlated confounders. The correlation is an estimate for the proportion of variance that is shared between two traits, based on common genetic factors.

It is therefore a useful 'hypothesis generating tool' that the correlation between two traits is due to pleiotropy or linkage. In the case of pleiotropy, the behaviour of one gene or genetic alteration may affect two distinct traits, or may affect one trait, but the expression of one trait has an effect on the expression of the second trait. Alternatively, the correlation may be caused by linkage if two distinct genes that affect one trait each, are in close proximity and thus in LD with one another. These two genes and thus the effect variants have on the two traits, will be inherited together and observed as a correlation. While numerous traits are likely to have a genetic correlation, it is not possible to conclude that the shared genetic variants have the same roles in the same mechanisms between the two traits. In addition, this particular method (cross-trait LD score regression) can be subject to bias from population stratification such as the impact of assertive mating on correlation estimates[255].

A trait representing persistent tinnitus had the greatest, significant genetic correlation with the *HDiff* phenotype, out of the traits that were tested (selected based on the publicly available data on LD hub). Therefore, this analysis provides a hypothesis that the two traits may be caused by common biological mechanisms. Tinnitus has been linked to hearing phenotypes in numerous cohorts yet a third of people with hearing loss do not have tinnitus. Understanding the relationship between these two traits could aid the development of either common diagnostics and treatments or the identification of distinct forms of either hearing loss or tinnitus. To date, only one GWAS has been performed on tinnitus independently of the multi-trait UKBB work by the Neale lab. The study was underpowered

(167 tinnitus cases and 749 controls) and thus no significant loci were identified yet suggestive associations allude to possible metabolic pathways that could be implicated in tinnitus pathology[302]. Identifying and understanding the genetic loci causal of the correlation observed between tinnitus and hearing loss in this genetic correlation analysis could allude to the pathology of subsets of tinnitus patients and thus strategies for managing either or both of the two traits. In order to achieve this, a high powered GWAS study of a robust tinnitus phenotype is required.

Having grouped the traits into five broad categories, pain, breathing difficulties and depressive symptoms all have relatively high correlations with hearing loss. Again, like tinnitus, depressive symptoms have previously been associated with hearing loss[195]; a higher incidence of depressive symptoms has been reported for hearing loss sufferers, in addition to some studies reporting a reduced depressive symptoms in hearing aid users[180,181]. These links have been discussed in detail in Chapter 1. The association between auditory function and pain has been observed before, both in studies of hyperacusis and central nervous system functioning. Studies have also linked specific pain disorders with hearing traits; in the HUNT study sample, those with fibromyalgia and other forms of musculoskeletal pain had an increased likelihood of reporting hearing loss (OR 4.578, 95% CI 3.622-5.787 and OR 4.523, 95% CI 3.077-6.647 in women and men), an association which held true in individuals that had been clinically diagnosed with fibromyalgia[303]. The authors hypothesise that this may be due to common dysregulation of the central nervous system that affects both pain response and hearing ability. Migraines have also been studied in connection with auditory traits and disorders of the vestibular system[303].

There may however be an underlying bias due to collinearity in these results. Many of the 41 traits that are highlighted here (significant p<, and a correlation of either <3 or >-3) are also based on self-report measures and thus an individual's perceived overall health scores may result in correlation between these self-report measures. In addition, the genetic correlation analysis presented here is restricted to traits that have publicly available GWAS summary statistics on LD hub[251] and the Bonferroni significance threshold determined by the number of traits included in the correlation analysis. While this is evidence of correlation between the traits that are tested in the samples that the original summary statistics are derived from, there are numerous other traits that are likely correlated and samples where these correlations may not be applicable. Currently, this method cannot be applied to recently admixed populations[251] and so association analysis of multiple traits is required in these populations, independent on these northern-European decent genetic correlations. Despite this, the current methods for genetic correlation analysis provide are

crucial tools for identifying pleiotropy between common complex traits and disease and thus enhance our understanding of individual traits and interactions between them.[304]

Chapter 5 - Association to function

5.1 Introduction

The use of GWAS to identify phenotype-SNP associations in complex traits has advanced rapidly over the past 10 years. Conversely, the translation of these SNP associations into useful knowledge of functional biological mechanisms has severely lagged behind. Figure 5.1 is from a review published in 2016 that demonstrates this point in terms of the number of publications; the number of functional follow-up studies from GWAS associations is a magnitude smaller than the number of GWAS published.

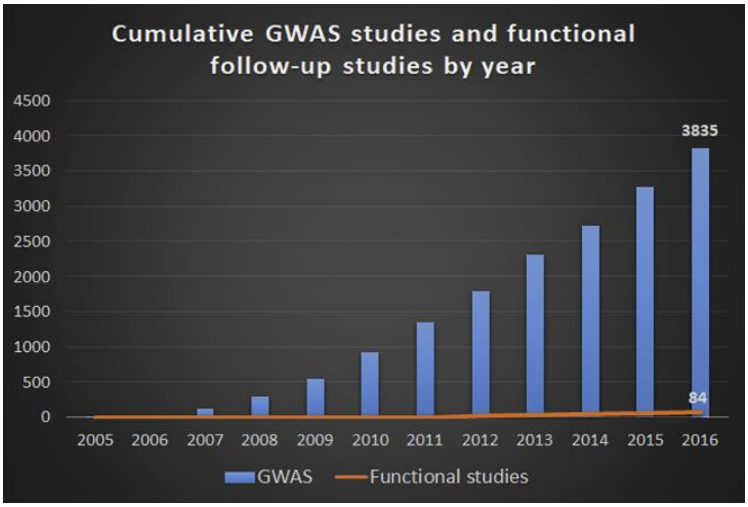

Figure 5.1. Bar chart displaying the cumulative number of GWAS studies compared to the cumulative number of functional follow-up studies by year.

The EBI GWAS catalog was used to determine the number of total GWASs reported from 2005 through the end of 2016, which are shown in blue. The number of post-GWAS functional studies reported each year were also identified (orange line) by (1) reviewing the titles, and in some cases, abstracts, of all research articles published in 23 relevant biomedical research journals∗, and (2) searching PubMed using the keywords "causal variant" or "functional variant." Additional studies were identified through references from primary research or review articles found as described. Figure and Legend from Gallagher, 2018[280].

Performing follow-up analysis of GWAS findings is the only way to convert SNP associations into meaningful, biological information that can be used to understand multiple traits and to even diagnose, treat and manage conditions. There are several statistical, bioinformatic and bench-based approaches that can be used to translate SNP associations to functional biological knowledge. In chapter 4, imputed SNP data were used to test for association with two phenotypes. Over 2,200 SNPs are significantly associated with either *HDiff* or *HAid*.

The initial post-GWAS steps are also presented in Chapter 4. Following association analysis, a statistical analysis approach called conditional (and sometimes joint) analysis was employed to isolate the lead SNP or SNPs at each locus to determine whether one or multiple variants are causing the association that is observed in each region[305]. In some cases, but not possible here, a region of association can then be fine mapped using via genome

sequencing, either of the region of the interest, or by using WGS or WES. Functional variants that were not included in the genotyped or imputed data can then also be tested for association[305].

The lead SNPs that were identified via conditional and joint analysis may be the functional variant at the locus or the SNP that best captures the variation. Replication analysis was then conducted with the aim to validate these findings. Only three of the lead SNP associations replicate at a Bonferroni-corrected level of significance. The relatively small size of the replication sample likely resulted in a lack of power to replicate a greater number of the associations. The work in chapter 4 also includes pathway analysis; a technique to identify gene and protein networks and thus hypothesise molecular and biological function of genes within associated regions. This method combines the effects of multiple genes to select networks and pathways that show significant enrichment with the phenotypes of interest.

In previous hearing GWAS, a small number of approaches have been used to translate association to function as described in the 'Genome-wide association studies' subsection of 1.4.1. The first ARHI GWAS (Friedman, 2009[109]) identified glutamate receptor, metabotropic 7 (*GRM7)* as the most highly associated gene candidate, performed histochemical studies in conjunction with the GWAS. Here just one SNP reached genome-wide significance in the study. It is in close proximity to the *GRM7* transcript and therefore *GRM7* was identified as the primary gene candidate. The histochemical data revealed mGluR7 (encoded by the mouse homologue of *GRM7*) expression in sensory hair cells and in spiral ganglion neurons. The third GWAS (Girotto, 2011[114]) included bioinformatic analysis in the form of an *in-silico* gene network construction. In the fourth study, (Nolan, 2013[92]) knock-out mice for the primary gene candidate, estrogen-related receptor gamma (*Esrrg*), were tested for relevant phenotypes and used for histological analysis, similar to the fifth study where mice were used to test for SIK family kinase 3 (Sik3) protein expression in cochlear tissue at different stages of development; a SNP in intron 6 of the *SIK3* transcript was the only genome-wide significant SNP identified[111]. Lastly, sequencing data were used to fine map regions of interest (Vuckovick, 2015[112]). The second (Van Laer, 2010[107]), seventh (Fransen, 2015[38]) and eighth (Hoffman, 2016[113]) ARHI GWAS were not published in conjunction with functional analysis.

Three of the studies performed immunohistochemistry in conjunction with the GWAS; the focus was on the localisation of murine proteins coded by mouse homologues of *GRM7*[109], *ESSRG*[92] and *SIK3*[111]. All three of these studies confirmed expression of the candidate protein in cochlear tissue. In addition, *Grm7* expression was observed in a cochlear tissue sample from an adult male (aged 83 with a PTA of 22dB), at the interdental cells of the spiral limbus,

IHCs, OHCs, Hensen's cells, type II fibrocytes of the spiral ligament and at the spiral ganglion neurons[109]. These three studies demonstrate the use of immunohistochemistry to translate an associated genetic variant to a hypothesised site of pathology within cochlear tissue.

For all previous ARHI GWAS it was clear which targets ought to be followed up; all studies that performed functional analysis detected only one significant locus or, in the absence of significance, selected the SNP with the highest association. Here, there are a high number of significant loci across the genome. Performing functional analysis on each of these candidates is not possible within this thesis due to practical and economical constraints. It is therefore necessary to prioritise a subset of candidates to investigate with functional analysis. Determining which gene is affected by the lead variant can be difficult; a number of the lead SNPs are not located within gene transcripts, and in gene dense regions, the associated SNPs may span multiple gene transcripts.

Additionally, in this study, due to the lack of power to validate associations by replication, prioritisation and selection of viable candidates for prioritisation is essential to warrant functional follow up analysis. To reduce bias in prioritisation and to not 'cherry pick' gene targets, where possible, a systemic evaluation ought to be followed for each of the loci. This framework is set out in 5.2.1 and is devised to prioritise gene candidates based on a number of criteria; the proximity of associated SNPs to the gene transcript, prior knowledge of gene function and function within the auditory system, strength of replication statistics, presence of a gene-specific mouse model, antibody availability and results from previous tissue-specific gene expression screens. This framework is used to select candidates for immunohistochemistry on mouse inner ear tissue samples. The aim of this immunohistochemistry study is to contribute to the validation of individual gene associations and hypothesise their putative function in the inner ear.

The concept of immunohistochemistry was first devised in the 1940s but it is still relevant for many current uses in research and diagnostics[306]. It exploits the binding properties of antibodies to specific protein domains (epitopes) and is means of detecting the presence of an epitope in a tissue sample. In this context, it can be utilised to determine whether a protein is expressed in the cochlea, and in which structure the protein is localised. The technique is of value in hearing research as the auditory system contains multiple structures that have distinct functions; there are over 26 different cell types in the cochlea alone. Each structure is highly specialised for specific functions and so determining the location of gene expression is a valuable means to explore putative functions.

The majority of the bench-based functional approaches used in auditory research use mice as the model organism. The location of the auditory system, inside the temporal bone and in brain tissue, means that it is impossible to access tissue samples from human patients

and, as the cochlea is encased in the temporal bone, near-impossible post-mortem. This impedes bench-based research and results in a lack of tissue-specific expression data on databases available for *in silico* analysis. Animal models are therefore a valuable resource in the study of the auditory system. As discussed in a subsection of chapter 1, 'Post-GWAS analysis', the mouse is the most common model used in auditory research due to the common anatomy, physiology and conserved genomic functions between mice and humans[238].

5.2 Results & Discussion

The genetic loci significantly associated with *HDiff* and *HAid* phenotypes (Chapter 4) can be used to hypothesise mechanisms involved in auditory function and pathology. The first aim of this chapter is to prioritise genes within these loci as potential candidates for functional analysis. All lead SNPs identified by conditional and joint analysis in chapter 4 are subset into groups and assessed for prioritisation. The second aim is to select (based on this framework) a subset of these candidates and perform immunohistochemistry in adult mouse cochlear samples.

The results of this chapter are presented in four sections, 5.2.1-5.2.4 (Figure 5.2). In section 5.2.1 each of the 48 lead SNPs identified by conditional analysis is reviewed based on genomic location and a primary (protein-coding) candidate gene from each locus was selected for further investigation. As immunohistochemistry is the selected method, the focus is on protein-coding genes (rather than genomic features such as regulatory elements). These primary gene candidates are then subset into two groups; (1) genes that already have an established role in hearing function and/or auditory pathologies and (2) those that are novel associations with hearing function and pathologies.

Section 5.2.2 is a summary and discussion of primary gene candidate genes that are known to have a role in auditory function or that cause hearing phenotypes. The aim of section 5.2.2 is to use previous knowledge of these known 'hearing genes' to speculate putative ARHI mechanisms. Section 5.2.3 is a report on the remaining the primary gene candidates that have not previously been linked to hearing function. The aim of section 5.2.3 is to assess each candidate gene for use in functional annotation, primarily for immunohistochemistry in the adult mouse, and also for future functional analysis. The results of the protein localisation analysis are presented in section 5.2.4.

As a full review of 48 loci is not possible to report within this chapter, key data sources are selected, and the relevant literature is summarised. A number of assumptions are made in order to subset and rank the SNPs and subsequent loci. While this is a necessary part of the prioritisation process, there are limitations with these methods that ought to be considered when reviewing the subsequent data such as a literature bias for well characterised genes or those that were included on RNAseq screens in auditory tissue. As the categorisation of genes in each section dictates the content of the following section (Figure 5.2) critical discussions of each of the four results sections are included alongside the results, instead of a combined discussion at the end of the chapter.

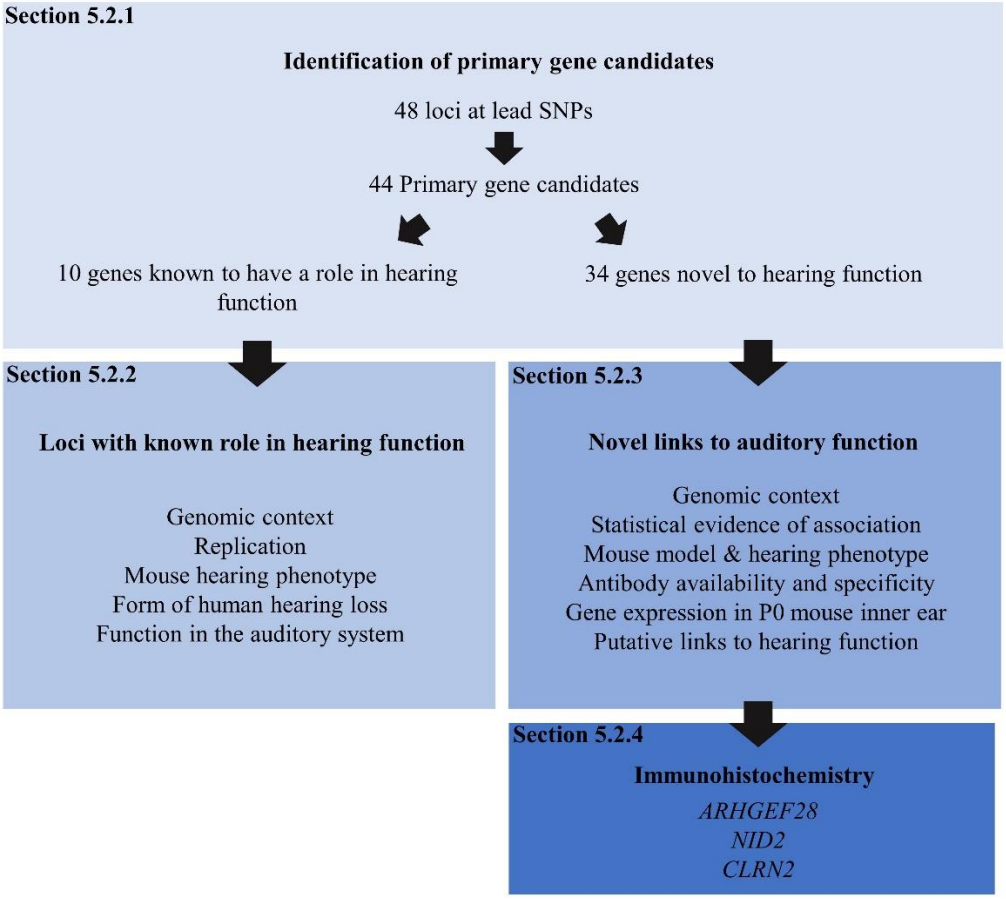

Figure 5.2. Workflow of results presented in Chapter 5.

The results are divided into four sections and presented in 5.2.1-5.2.4. In section 5.2.2 and Section 5.2.3, each of the primary gene candidates is assessed based on the criteria listed in the corresponding boxes in this figure.

5.2.1 Identification of primary gene candidates from *HDiff* and *HAid* genetic association analysis

In this section primary gene candidates at each of the associated regions identified in the *HDiff* and *HAid* GWAS (presented in chapter 4) are selected and evaluated. At each locus a primary gene candidate is selected based on the proximity of the gene transcript to the lead SNP at the locus. Each locus is also assessed based on whether additional gene transcripts within the locus contained SNPs that were (1) in high LD (r^2>0.6) with the lead SNP and (2) were significantly associated (P<E5-08) with the phenotype.

The higher the number of gene transcripts that contain significant SNPs in the LD with the lead SNP, the lower the confidence in gene candidate selection. Conversely, at loci where only one gene transcript contains significant SNPs that are in high LD with the lead SNP, there is more confidence in the selection of the gene candidate. Locus plots that visualise each significant lead SNP and the flanking 1-200kb genome region, are included in the

Appendix, locus plots (a-vv). The LDlink tool[257] is used to test for LD scores between a lead SNPs and genome-wide significant SNPs in the same region see - Methods 2.4.1.

For a number of the loci, more than one gene at the locus contains genome-wide significant SNPs that are in high LD with the lead SNP. Therefore, at these loci it is difficult to conclude which gene is the primary candidate. Here the primary candidate is selected based on either the proximity of the gene to the lead SNP (*ZNF318, TYR, ATP11B, AGO2, ADK* and *SNAP91),* prior knowledge regarding a known gene role in hearing function (*CLRN2*, based on knowledge of the role of clarin-1 in USH3 as reported in Section 1.2.1) or the presence of a second lead SNP in the same gene (*ARHGEF28*). While these assumptions aid in the selection of candidates to study, other genes in proximity ought to be considered in further analysis and when interpreting results from functional analysis.

Following the selection of a primary gene candidate at each locus, each gene is evaluated with respect to whether it has previously been linked to a hearing pathology and, or has an established role in auditory function (Figure 5.2). To determine this, list of gene candidates is cross referenced with the published list of genes reported to underlie deafness in either humans or mice (2018)[260], and with any report when using the PubMed search terms '<gene name>, hearing' and '<gene name>, deafness'. To present these data, the lead SNPs are divided into three tables based on whether they are in exon regions; Table 5.1, intronic; Table 5.2, or intergenic; Table 5.3.

5.2.1.1 Results & Discussion: Identification of primary gene candidates from *HDiff* and *HAid* genetic association analysis

Table listing all lead SNPs from *HDiff* that are in gene exon regions							
Lead SNP	**Pheno**	**P-value**	**No.**	**Gene names**	**Primary candidate**	**Aud. Func.**	**Variant consequence (REVEL score)**
rs36062310	*HDiff*	1.90E-22	1	*KLHDC7B*	*KLHDC7B*	-	Missense var. (0.088)
rs6453022	*HDiff*	1.70E-21	1	*ARHGEF28*	*ARHGEF28*	-	Missense var. (0.055)
rs9493627	*HDiff*	1.40E-13	1	*EYA4*	*EYA4*	✓	Missense var. (0.424)
rs5756795	*HDiff*	5.10E-12	1	*TRIOBP*	*TRIOBP*	✓	Missense var. (0.102) / 3' UTR variant / NMD target
rs143282422	*HDiff*	2.40E-08	1	*CDH23*	*CDH23*	✓	Missense var. (0.198) / non-coding var.
rs12552	*HDiff*	4.80E-08	1	*OLFM4*	*OLFM4*	-	3' UTR variant

Table 5.1. Lead SNPs from *HDiff* association analysis that reside in gene exons.

Lead SNP; lead SNP in joint and conditional analysis performed in chapter 4, Pheno; phenotype used for GWAS, P-value; BOLT-LMM infinitesimal mixed model association test p-value of the lead SNP in GWAS presented in chapter 4, No.; the number of genes at the locus that contained SNPs that had a genome-wide significant p-value and that are in high LD (r^2>0.6) with the lead SNP, Gene names; gene names for the genes totalled in 'No.', Genomic position; genomic position of the lead SNP, Primary candidate; gene name of the nearest protein-coding gene to the position of the lead SNP, Aud. Func.; ✓ indicates that the proximity-based primary candidate has an established role in auditory function, Variant consequence; the VEP predicted consequence of the variant at the lead SNP position, NMD; non-mediated decay, UTR; untranslated region, NC; non-coding. REVEL prediction values for missense mutations are noted in brackets, scores range from 0-1 with higher scores reflecting greater likelihood that the variant is disease-causing[307]. As rs12552 is not a missense variant, no REVEL score is listed.

Table listing all lead SNPs from *HDiff* and *HAid* that are in gene intron regions

Lead SNP	Pheno	P-value	No.	Gene names	Primary candidate	Aud. Func.	Variant consequence
rs4597943	*HAid*	2.10E-11	1	*ARHGEF28*	*ARHGEF28*	-	Intron var. / Regulatory region var.
rs759016271*	*HDiff*	6.10E-21	2	*CRIP3, ZNF318*	*ZNF318*	-	Intron var. / NMD target / NC transcript
rs9321402	*HAid*	3.00E-10	1	*EYA4*	*EYA4*	✓	Intron var. / NC transcript
rs132929	*HDiff*	2.20E-13	1	*BAIAP2L2*	*BAIAP2L2*	✓	Intron var. / NC transcript exon var.
rs1566129	*HDiff*	1.40E-11	1	*NID2*	*NID2*	-	Intron var.
rs1566129	*HAid*	2.50E-09	1	*NID2*	*NID2*	-	Intron var. / upstream gene var.
rs12225399	*HDiff*	8.60E-11	1	*PHLDB1*	*PHLDB1*	-	Intron var., NMD transcript var., NC transcript var., upstream gene var., regulatory region var.
rs55635402	*HDiff*	2.90E-10	1	*TUB*	*TUB*	✓	Intron var., downstream gene var., regulatory region var.
rs9677089	*HAid*	2.00E-10	1	*SPTBN1*	*SPTBN1*	✓	Intron var.
rs62033400	*HDiff*	2.90E-10	1	*FTO*	*FTO*	-	Intron var., NMD target, NC transcript var.
rs13277721	*HDiff*	3.30E-10	2	*AGO2, PTK2*	*AGO2*	-	Intron var., NC transcript var., NMD target
rs2236401	*HDiff*	9.30E-10	1	*SYNJ2*	*SYNJ2*	✓	Intron var.
rs4947828	*HDiff*	1.00E-09	1	*GRB10*	*GRB10*	-	Intron var., NC transcript var.
rs6597883	*HDiff*	1.00E-09	1	*CTBP2*	*CTBP2*	✓	Intron var., NC transcript var.
rs10901863	*HAid*	2.60E-08	1	*CTBP2*	*CTBP2*	✓	Intron var., NC transcript var.
rs34442808	*HDiff*	1.30E-09	2	*MCTP1 & ANKRD32*	*MCTP1*	-	Intron var., NC transcript var.
rs835267	*HDiff*	1.60E-09	1	*EXOC6*	*EXOC6*	-	Intron var., NMD target, regulatory region var.
rs4948502	*HDiff*	1.70E-09	1	*ARID5B*	*ARID5B*	-	Intron var.
rs10824108	*HDiff*	3.00E-09	3	*VCL, AP3M1, ADK*	*ADK*	-	Intron var., NC transcript var.
rs217289	*HDiff*	4.90E-09	1	*SNAP91*	*SNAP91*	-	Intron var., NMD target
rs3915060	*HAid*	9.70E-09	1	*ILDR1*	*ILDR1*	✓	Intron var., NC transcript var.
rs17671352	*HDiff*	1.00E-08	6	*DLG4, ACADVL, DVL2, PHF23, CTDNEP1*	*ACADVL*	-	Down & upstream gene var., splice region var., intron var., NC transcript exon var., NMD target, regulatory region var.

Table listing all lead SNPs from *HDiff* and *HAid* that are in gene intron regions

Lead SNP	Pheno	P-value	No.	Gene names	Primary candidate	Aud. Func.	Variant consequence
rs12938775	*HDiff*	1.60E-08	1	*PAFAH1B1*	*PAFAH1B1*	-	Intron var., NC transcript var.
rs76837345	*HDiff*	1.90E-08	1	*CHMP4C*	*CHMP4C*	-	Intron var.
rs9366417	*HDiff*	2.10E-08	1	*SOX4*	*SOX4*	-	Intron var., NC transcript var.
rs9691831	*HDiff*	3.10E-08	1	*TMEM213*	*TMEM213*	-	Intron var.
rs141403654	*HDiff*	3.50E-08	1	*AGBL2*	*AGBL2*	-	Intron var.
rs10927035	*HDiff*	4.90E-08	1	*AKT3*	*AKT3*	-	Intron var.

Table 5.2. Lead SNPs from *HDiff* and *HAid* association analyses that reside in gene introns. Legend as in Table 5.1.

Table listing all lead SNPs from *HDiff* and *HAid* that are in intergenic regions

Lead SNP	Pheno	P-value	No.	Gene names	Gen pos.	Primary candidate	Aud. Func.	Variant consequence
rs6890164	*HDiff*	3.30E-19	3	*ARHGEF28, ANKRA2, UTP15*	6177bp up	*ARHGEF28*	-	Intergenic var.
rs7951935	*HDiff*	7.80E-17	2	*TYR, NOX4*	1472bp up	*TYR*	-	Intergenic var.
rs35186928	*HDiff*	1.70E-15	0	-	13352bp up	*HLA-DQA1*	-	Intergenic var., regulatory region var.
rs35414371	*HDiff*	1.60E-11	2	*QDPR, CLRN2*	1965bp down	*CLRN2*	-	Downstream gene var.
3:182069497_TA_T*	*HDiff*	4.10E-11	0	*ATP11B*	441791bp	*ATP11B*	-	-
rs62188635	*HDiff*	4.70E-10	1	*KLF7*	50519bp up	*KLF7*	-	Intron var. in NC transcript
rs12027345	*HDiff*	3.60E-09	1	*MAST2*	12668bp up	*MAST2*	-	Intergenic var.
rs13093972	*HDiff*	5.50E-09	1	*ZBTB20*	121137bp up	*ZBTB20*	-	Intron var. in NC transcript
rs62015206	*HDiff*	7.70E-09	2	*MAPK6, BCL2L10*	15613bp down	*MAPK6*	-	Upstream gene var., intron NC var.
rs10475169	*HDiff*	9.30E-09	0	-	190445bp down	*IRX2*	-	Intergenic var.
rs7525101	*HDiff*	1.50E-08	0	*LMX1A*	61973bp up	*LMX1A*	✓	Intergenic var.
rs3890736	*HDiff*	2.20E-08	1	*GFRA2*	15676bp down	*GFRA2*	-	Intergenic var., regulatory region var.
rs7823971	*HAid*	2.70E-08	1	*RP11-1102P16.1*	-	*RP11-1102P16.1*	-	Intron var. in NC transcript
rs4611552	*HDiff*	3.60E-08	1	*CCD68*	9362bp up	*CCD68*	-	Intergenic var., regulatory region var.

Table 5.3. Lead SNPs from *HDiff* and *HAid* association analyses that reside in intergenic regions.

Table legend as in Table **5.1**. *denotes that the SNP is not in the Locuszoom reference and so the locus plot was created using a SNP (rs11280821) that in close proximity to 3:182069497_TA_T.

This section evaluates 44 primary gene candidates for potential use in functional analysis. The results include an assessment of the confidence of the gene selections, based on the genomic context of the lead SNP. 34 of the lead SNPs reside within the gene transcript of the selected primary gene candidate, 6 in exon regions (Table 5.1) and 28 in intron regions (Table 5.2). For the majority of the primary gene candidates, the lead SNP resides in either an intron or exon of the gene and no other genes at the locus contain genome-wide significant SNPs in high LD with the lead SNP. This provides good confidence of the gene selection at these loci. However, 14 lead SNPs are in intergenic regions (Table 5.3) and 7 of the loci have genome-wide significant SNPs in high LD with the lead SNP, in more than one gene (Table 5.1, Table 5.2, Table 5.3 column titled 'No.' to represent the number of possible *cis* gene candidates under these assumptions). In these cases, there is greater ambiguity and thus less confidence when selecting the primary gene candidate. For example the region flanking the lead SNP rs62015206 in Figure 5.3 below, displays a region where a cluster of SNPs in high and in moderate LD with the lead SNP covers a region containing multiple gene transcripts. Here, the primary candidate is *MAPK6*, due to the SNPs in high LD (coloured in red) with the lead SNP (purple) are in the *MAPK6* transcript.

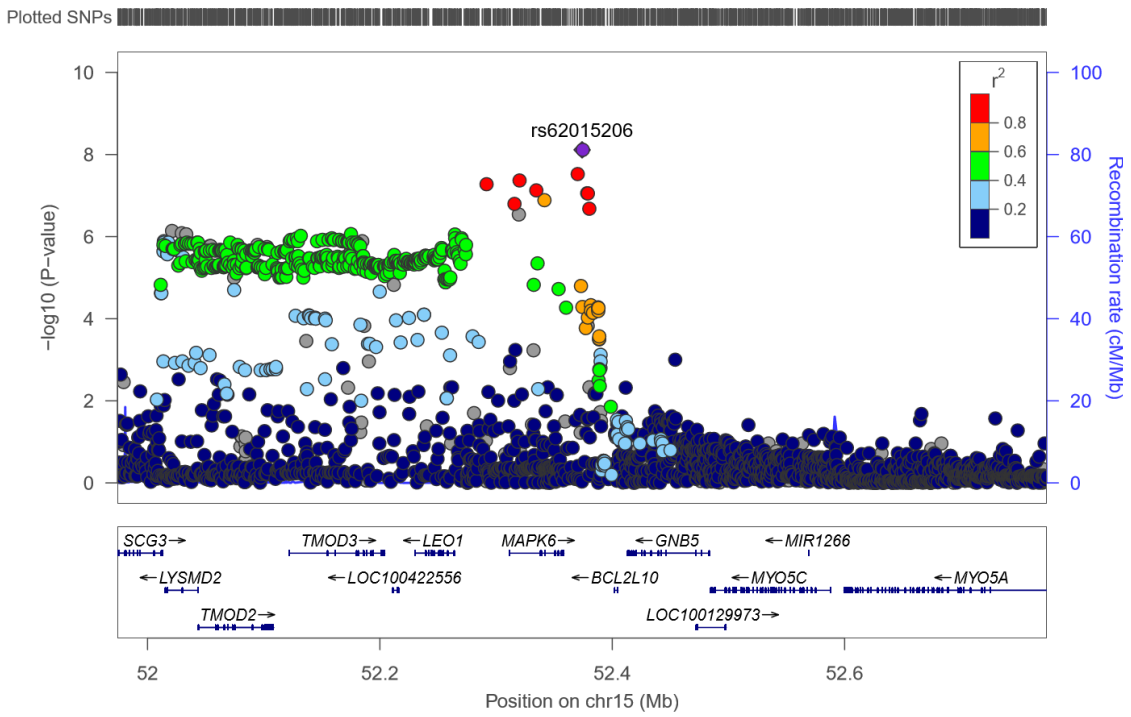

Figure 5.3. Locus plot displaying the genome region flanking the lead SNP rs62015206

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey.

Secondly, the results include data from Ensembl Variant Effect Predictor (VEP), which predict the functional consequence of each variant. Five of the six lead SNPs that reside in exon regions are predicted to result in missense mutations. A missense mutation leads to

an amino acid substitution. REVEL scores which estimate the deleteriousness of such a change[307], predict that the SNP rs9493627 (within an *EYA4* exon region, REVEL score 0.424), has the greatest likelihood, of the five, to be disease-causing.

While a number of the lead SNPs may be the causal variant, it is likely that the variants listed are not the causal variant but are in high LD with the causal variant. However, this information can be used to identify possible regulatory regions and hypothesise variant consequences related to gene function. Most of the variants have multiple predicted consequences. This is largely due to the different transcripts and splice forms of genes in close proximity to the variant, but from this association analysis alone, it is not possible to conclude which transcript(s) relate to the variant association and thus which consequence is likely to be causal of the association with the phenotype.

Thirdly, the results include a subset of 10 genes that are known to have a role in hearing function or pathology in either mouse or human subjects. Current knowledge of the function of these genes may reveal previously unknown ARHI mechanisms, that are already known to cause other forms of hearing impairment. These findings could also challenge the current knowledge of these specific genes. For example, a gene newly associated with ARHI that was previously known to cause an early-onset congenital defect (and thus had been determined as a developmental gene) could now be hypothesised to have a post-developmental role such as maintenance or stress response. In addition, the fact that almost a quarter of these gene candidates have previous links to hearing, can be interpreted as a validation of the association results, for using phenotypes derived from self-report measures. The loci containing known hearing genes are explored further in the next section (5.2.2).

Primary gene candidates are selected based on the proximity to lead SNPs rather than any known genomic functions present within the region such as regulatory variants. While it necessary to focus on protein coding genes at each locus, especially when selecting candidates for immunohistochemistry, it is important to consider the fact that the true functional target gene or variant represented by the association may reside elsewhere at the locus or even at a different locus by either cis or trans interactions. To address this in future work, primary gene candidates could be determined using a combination of their proximity to lead SNPs and biological functionality and thus mapped to genes based on positional, eQTL and chromatin interaction data, possible with platforms such as FUMA[233]. For example, the lead SNP rs9366417 identified in *HDiff* is here reported to be in close proximity to the gene *SOX4*, but the SNP lies in the lncRNA transcript *CASC15*, which may be the true functional target represented by the variant association. A locus plot corresponding to this region is displayed in a following section, in Figure 5.6.

5.2.2 Loci associated in *HDiff* and *HAid* genetic association analysis that contain genes with a known role in hearing function or hearing conditions

Ten of the 44 loci identified in Chapter 4 have previously been linked to auditory function (Table 5.1, Table 5.2, Table 5.3, column 9). The importance of some these findings is briefly discussed in a previous section; section 4.3 'Findings from genetic association analysis that are linked with forms of congenital deafness'. All ten loci contain genes that either (1) code for proteins that have an established role in the auditory pathway or (2) which have been implicated in hearing disorders and conditions. Only one locus (lead SNP rs5756795 in *HDiff* analysis) contains a gene that has been linked to AHRI; *TRIOBP*[113]. In this section each of these previous links to hearing loss are discussed, and the findings are discussed at the end of the section.

5.2.2.1 Gene Function summaries

EYA4 (DFNA10)

Both *HDiff* and *HAid* analyses have lead SNPs that lie in the *EYA4* gene; rs9493627 in exon 10 (*HDiff*, p=1.40E-13) and rs9321402 in intron 10 (*HAid*, p=3.00E-10). Neither of the two SNPs replicated in the replication meta-analysis, yet variants in *EYA4* have previously been linked to autosomal dominant non-syndromic hearing loss[204,288,289]. The EYA protein family is a class of protein tyrosine phosphatases and different EYA family members have distinct links to a number of conditions and diseases including multiple forms of hearing loss[308].

EYA4 specifically, codes for Transcriptional Coactivator and Phosphatase 4 and is located at the DFNA10 locus. The first direct link between *EYA4* and human auditory function was identified in two families that present cases of late-onset hearing loss[309]. Since this initial finding, multiple other families with autosomal dominant non-syndromic hearing loss have been found to have insertions and variants in the *EYA4* gene. The affected region differs between the families, but all variants result in premature stop codons in exon 14, exon 20, exon 13and intron 14[289,310].

The expression profile of *EYA4* in the inner ear has been shown to differ between animals and it's been hypothesised that this may reflect the different hearing phenotypes observed between mice and humans. In *Eya4* mutant mice, hearing loss is early in onset and is caused by otitis media; Eya4$^{-/-}$ mice exhibit a phenotype indicative of middle ear effusion with tympanic membrane retraction. This is suggestive of *Eya4* expression and function in the middle ear[311]. This is in contrast to findings in non-human primates (marmosets), where *EYA4* expression in adult inner ears is found in the hair cells, spiral ganglion and a range of supporting cells. The observed phenotype resulting from *EYA4* mutations in humans is also different from that in rodents; diagnosed patients exhibit sensorineural hearing loss[204].

From association analysis alone, it is not possible to confirm whether the pathology in the UKBB sample lies in the middle or inner ear. However, an inner ear pathology can be hypothesised. This is based on the observation of sensorineural defects in primate studies of *EYA4* function, and because the GWAS the phenotypes were derived as surrogate measures for ARHI, a form of sensorineural hearing loss.

SPTBN1

The lead SNP (rs9677089, p=2.00E-10) resides in intron 2 of the *SPTBN1* gene, which codes for βII Spectrin, a member of the beta-spectrin gene family. β II-Spectrin has been identified in the lateral wall of outer hair cells[312] and at the apical region of inner hair cells in mouse cochlear tissue[313]. The most recent study on spectrin function in cochlear hair cells, found that β II-Spectrin, along with α II-Spectrin, provides mechanical support for HC rootlets. The work also revealed that βII spectrin was not required in the early development of HC stereocilia in mice; the spectrin rings form ~P14, at the time when mice acquire the ability to hear and are thus likely to have a role in HC function or maintenance. In aged mice, the authors found levels of spectrin ring disruption at OHCs correlates with ABR threshold deterioration[296]. In addition, Atoh1-Sptbn1$^{-/-}$ mice (lacking βII-spectrin in inner ear HCs by mating Atoh1-*cre* and Sptbn1$^{flox/flox}$ mice) display a severe deafness[296]. Despite this established role in the cochlea, prior to this study no link had been made between genetic variants in *SPTBN1* and human hearing pathologies. Based on the previous work in mice, it could be hypothesised that the *HDiff* phenotype may in part be caused by defects in the support of HC rootlets.

BAIAP2L2

The lead SNP rs132929 (p=2.20E-13) in *HDiff* analysis is located in intron 7 of the *BAIAP2L2* gene which codes for Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 2. *Baiap2l2* was a gene identified in the recent large-scale mutant mouse screen phenotypes[243]; *BAIAP2L2* mutant mice were observed to have a mild hearing loss at all frequencies. The *BAIAP2L2* gene transcript is also in a locus that is implicated in some cases of Waardenburg syndrome (WS). A WS patient with a full deletion of the locus (including the full *BAIAP2L2* transcript), presents with profound bilateral sensorineural hearing loss[290]. Elsewhere, there are few reports of the general function and pathological impact of *BAIAP2L2*. These findings are evidence of BAIAP2L2 having a role in hearing function but do not reveal a biological cause for BAIAP2L2-linked pathology and thus do not provide an insight into the pathological mechanisms underlying the *HDiff* association at this locus.

TRIOBP

The SNP rs5756795 had an association p-value of 5.10E-12 in the *HDiff* discovery analysis and lies in an exon of the *TRIOBP* gene transcript, which codes for TRIO and F-Actin Binding Protein. *TRIOBP* is the only gene that contains a lead SNP from either *HDiff* and *HAid* and that has previously been linked to ARHI. The previous link was also in a GWAS; a SNP in close proximity to the *TRIOBP* gene transcript reached genome-wide significance in the Hoffman 2016 ARHI GWAS[113]. The significant SNP in the Hoffman analysis is in high LD (r^2=0.96) with a missense mutation in the *TRIOBP* transcript and is in complete LD with the lead SNP in *HDiff* analysis, rs5756795.

In addition to ARHI, mutations in *TRIOBP* are known to cause recessive non-syndromic hearing loss DFNB28[53] in 9 Palestinian families. 27 members from 7 of the families were shown to have homozygous nonsense mutations in the SH3 binding domain of *TRIOBP*, while 3 members from the remaining two families exhibited heterozygous mutations for two nonsense mutations. Following this finding, localisation of the protein products from different *Triobp* splice forms were studied in the mouse cochlea. Localisation was identified in inner ear hair cells, and localised with F-actin along the length of the stereocilia[53]. All previous *TRIOBP* variants that have been linked to familial cases of hearing loss are located in exon 6 of the isoforms *TRIOBP4* or *TRIOBP5*[314]. The SNP that is here associated with *HDiff* lies in exon 7, and so may have a different impact on gene function at the stereocilia when compared to the functional effect of variants located in exon 6.

TUB

The lead SNP rs55635402 in *HDiff* analysis (p=2.90E-10) is located in an intron of the gene *TUB,* that codes for Tubby Bipartite Transcription Factor. This result provides the first evidence of a link between *TUB* and human hearing function. The *Tub* gene was initially studied in a mouse model of Usher Syndrome Type 1 (*USH1*) because Tub^{tub} mice exhibit a progressive hearing loss and retinal degeneration in the absence of further neurological phenotypes as observed in USH1[298]. Tubby-like proteins are a unique family of bipartite transcription factors[315], but the precise function of TUB in the mouse cochlea is not understood. It has since been observed that $^{tub/tub}$C57 mice display an accelerated OHC and spiral ganglion cell loss at the basal end of the cochlea relative to $^{tub/+}$ and $^{+/+}$ C57 mice[316]. The hypothesised function of TUB in the auditory pathway is a modifying effect at synaptic junctions between sensory cells and neurons[317]. The association with *HDiff* therefore may be suggestive of pathologies that are present at synaptic junctions.

SYNJ2

The lead SNP at the *SYNJ2* locus in the *HDiff* analysis, rs2236401 (p=9.30E-10), resides in an intron region of *SYNJ2. SYNJ2* codes for Synaptojanin 2, an inositol phosphatase that is known to function in recycling neurotransmitter vesicles[295]. This work provides the first link between *SYNJ2* and a hearing pathology in humans. The finding linking *SYNJ2* and hearing function was made in a mutant mouse mozart which displayed a progressive hearing loss caused by the deterioration of sensory cells, resulting in complete loss of hair cells[318]. This is suggestive of a role for *Synj2* in hair cell maintenance and function and therefore a possible site of pathology in *HDiff* cases.

Although not previously linked to hearing in a population cohort, *SYNJ2* has also been associated with longevity and cognitive function in the elderly[319]; a trait closely related to ARHI. Further to this, a study in 2012 reported an association with *SYNJ2* and cognitive abilities. Conserved longevity genes between yeast and nematode species[320] were used as candidate genes to test for an association with cognitive ability in a human population meta-analysis.

ILDR1 (*DFNB42*)

In the *HAid* analysis, the lead SNP rs3915060 (p=9.70E-09) lies in an intron of *ILDR1*, which codes for immunoglobulin-like domain-containing receptor 1. Variants in *ILDR1* have been associated with multiple cases of autosomal recessive non-syndromic and syndromic sensorineural hearing loss[321–324]. Based on this, *ILDR1* was also considered a candidate for common hearing impairment in the candidate gene analysis of the 2016 Hoffman study[113], where common variants with in *ILDR1* were found to be associated with the phenotype.

The function of *Ildr1* in the auditory system has been studied in mice; Ildr1$^{(-/-)}$ mice display a progressive degeneration of OHCs, and disruption to the tunnel of Corti is observed by p21[325]. Two further studies have identified roles for *Ildr1* in tricellular tight junction formation[326,327]. In addition, protein up-regulation (fold change>1.5) and down-regulation (fold change<0.5) was shown to differ in over 800 proteins, between Ildr1$^{(+/-)}$ and Ildr1$^{(-/-)}$ mice at P21[325]. This indicates that *ILDR1* may be a master regulator and, or have an upstream role in pathways in which these proteins are involved. The phenotype observed may therefore be caused by the modified expression of downstream targets.

LMX1A

The lead SNP from the *HDiff* analysis at the *LMX1A* locus, rs7525101 (p=1.50E-08) is in close proximity to the *LMX1A* gene transcript that codes for Lim homeobox transcription factor 1. Mutations in *LMX1A* were discovered to cause to progressive sensorineural hearing loss in two Dutch families, and to severe cochlear and vestibular defects in mouse mutants(mtl

and bsd)[328]. In the two families, the phenotypes are varied in the age of onset, rate of progression and whether vestibular defects are observed. There are no apparent developmental pathologies however; the hearing loss is consistent between the two families in that it is progressive.

The authors hypothesise that one copy of a functional *LMX1A* gene is sufficient for development of auditory and vestibular structures but not necessarily for the maintenance of function[328]. Although the lead SNP rs7525101 does not reside within the *LMX1A* transcript, the previous links to hearing phenotypes increase the likelihood that the target of the variant at this locus is in the function of *LMX1A*, either via a *cis* regulatory element or a SNP within the gene transcript that is in LD with this lead SNP.

CDH23

The lead SNP rs143282422, from *HDiff* analysis, lies in an exon of the *CDH23* gene transcript which codes for cadherin-23, a protein that forms the majority of the stereocilia tip-link structure at the inner hair cells of the organ of Corti[329]. Multiple mouse strains have point mutations at the Cdh23 locus, which cause a variety of hearing phenotypes. A key example is the *Cdh23^{ahl}* allele in the C57BL/6J strain, which is the most commonly used strain for transgenic analysis. The mutation causes an accelerated HL, with a progressive loss of inner and outer HCs and loss of the SG cells at the basal turn of the cochlea by seven months of age[238,241]. The locus containing Cdh23 is termed Ahl1 as it was one of the first loci found to be associated with ARHI in mice. The association with *HDiff* at this locus is the first link between *CDH23* and ARHI to be found in a human population, however methylation levels at the *CDH23* locus have been shown to have an association with ARHI risk in a sample of elderly women[330].

Homozygous mutations in *CDH23* are known to cause Usher Syndrome Type 1D, symptoms of which include profound congenital sensorineural hearing loss vestibular areflexia, and adolescent-onset retinitis pigmentosa. It has also been identified in cases of non-syndromic hearing loss[331,332] and is a candidate for common hearing impairment or presbycusis[330]. *CDH23* was recently associated with non-syndromic hearing loss in a meta-analysis of four Asian population samples at the p.P240L variant[333]. The association between *HDiff* and *CDH23* again suggests that the stereocilia may be some of the structures that cause *HDiff* when damaged or degenerated. The association with *HDiff* at this locus is therefore in part validated, as *CDH23* is known to cause other forms of deafness (such as USH1D), along with the hearing defects observed in the C57BL/6J mouse strain.

CTBP2

The lead SNP rs6597883 from *HDiff* lies in an intron of *CTBP2* which codes for the c-terminal binding protein 2 and its' alternative splice form that codes for ribeye, a major component of the hair cell ribbon synapse[294,334]. Anti-ctbp2 antibodies are commonly used as a marker for the ribbon synapse in protein localisation assays[113]. Although ribeye is an established component of the ribbon synapse, mutations in the coding transcript have not previously been associated with pathology in human hearing loss. At the ribbon synapse, multiple ribeye subunits interact via five distinct sites at A and B domains. Within the cochlea, these synapses are the site of rapid exocytosis of synaptic vesicles and thus have a specialised role in 'speed vesicle trafficking'[334]. Ribeye knock-down mice display a lack of IHC synaptic ribbons and diminished organisation of the IHC active zone topography (vesicle replenishment and Ca^{2+} channel regulation)[335]. These findings are the first to link *CTBP2* (or ribeye) to a human pathology, and the established role of ribeye in the ribbon synapse is further evidence that suggests *HDiff* could be caused by a synaptopathy.

Table 5.4 is a summary of the findings presented above, regarding the subset of gene candidates that have previous links to hearing loss or an established function in the auditory system. The legend for Table 5.4, which is on the following page, is:

Table 5.4. Primary gene candidates with established links to hearing function and pathologies.

Lead SNP; lead SNP in joint and conditional analysis performed in chapter 4, Geno position; genomic position of the lead SNP, Pheno; phenotype used for GWAS, Rep; ü indicates a nominal replication (P<0.05) in the replication meta-analysis presented in chapter 4, Known hearing gene; name of gene at locus that has an established link to hearing function or pathology, Mouse model with hearing phenotype; ü indicates that the mutant mouse model for the corresponding gene displays a hearing phenotype, Human HL; ü indicates that a variant in this gene has been linked to human hearing loss, Type of hearing loss; the type of hearing loss that individuals present due to a mutation in the corresponding gene, Known function; understood function of the corresponding gene in the auditory system.

Lead SNP	Geno position	Pheno	Rep	Gene	Mouse pheno	Human HL	Type of human hearing loss	Known function
rs9493627	Exon	*HDiff*	-	*EYA4*	✓	✓	Autosomal dominant non-syndromic hearing loss[204,288,289]	Tympanic membrane hypervascularity, retraction of the tympanic membrane, and middle ear effusions[336]
rs9321402	Intron	*HAid*	-					
rs9677089	Intron	*HAid*	-	*SPTBN1*	✓	-	Novel	Anchoring complexes of HC stereocilia[296]
rs132929	Intron	*HDiff*	-	*BAIAP2L2*	✓	✓	Syndromic sensorineural hearing loss[290]	-
rs5756795	Exon	*HDiff*	-	*TRIOBP*	✓	✓	Non-syndromic recessive[16] & ARHI[113]	Provides durability and rigidity for normal mechanosensitivity of stereocilia[53]
rs55635402	Intron	*HDiff*	✓	*TUB*	✓	-	Novel	Neuronal cell synaptic function, via interactions with MTAP1A[337]
rs2236401	Intron	*HDiff*	-	*SYNJ2*	✓	-	Novel	Hair cell maintenance possibly via degradation of phosphoinositide signalling molecules[318]
rs3915060	Intron	*HAid*	-	*ILDR1*	✓	✓	Recessive Non-syndromic Deafness, ARHI[113]	OHC maintenance and function, possibly via modulating tricellulin expression[338]
rs7525101	Intergenic	*HDiff*	-	*LMX1A*	✓	✓	Non-syndromic recessive hearing impairment[292]	Formation of endolymphatic duct and HCs in the basal cochlear region, and interaction with LMO in cochlear structure formation[339]
rs143282422	Exon	*HDiff*	✓	*CDH23*	✓	✓	Syndromic Autosomal recessive[291]	Cohesion of HC stereocilia with MYO7A and harmonin b[55], and tip link function with PCDH15[340]
rs6597883	Intron	*HDiff*	✓	*CTBP2*	✓	-	Novel	Component of the synaptic ribbon[341]
rs10901863	Intron	*HAid*	-					

Table 5.4 Primary gene candidates with established links to hearing function and pathologies.

5.2.2.2 Discussion: Loci associated in *HDiff* and *HAid* genetic association analysis that contain genes with a known role in hearing function or hearing conditions

Section 5.2.2 is a study into the subset of gene candidates that are already established as 'hearing' genes; the genes that either have a known role in auditory function or that have been linked to a hearing pathology. For four of the genes; *SPTBN1*, *TUB*, *SYNJ2* and *CTBP2,* this is the first report of evidence of a link to an auditory pathology in humans. Variants in each of these four genes are linked to hearing phenotypes in mutant mice, demonstrating conserved auditory functions between the two organisms and supporting the use of mice to identify pathogenic variants for human hearing function.

Of the ten genes, only two have been linked to forms of common hearing loss; *TRIOBP* and *ILDR1* were associated with common hearing impairment in a previous ARHI GWAS[113]. The remaining four genes that have previously been linked to human hearing loss; *EYA4*, *BAIAP2L2*, *LMX1A* and *CDH23*, cause rare or familial cases of severe, early or late-onset hearing impairment. As variants in these genes are here associated with a common hearing loss, the variants causing this pathology could be less deleterious than those observed in previous severe and rare cases. An example could be deleterious variants in a gene that codes for damage repair or damage resistance. If these variants resulted in an impeded or inhibited expression or function of the encoded protein, over time crucial structures may not be amply repaired, resulting in a progressive phenotype.

Further to this, for several of the genes discussed in this section, the region harbouring the pathogenic variant in *HDiff* or *HAid* differs between the previous reported cases. This may reflect the difference between the phenotypes observed; rare, severe cases in contrast to common variants causing a common impairment as assessed here. The location of a variant within or in proximity to a gene transcript can affect the severity and onset of the phenotype. Distinct amino acid substitutions in Connexin 26 (*GJB2*) cause different functional changes which are correlated with the severity of some forms of congenital, non-progressive sensorineural hearing impairment[43].

For example, pathogenic regions identified in the *EYA4* transcript that presented as rare, severe cases have been variants that produce premature stop codons in exons 14[289], exon 20[289], exon 13[289] and intron 14[310]. In the association analysis here, lead SNPs are identified in intron 1 (*HAid*) and exon 1 (*HDiff*) in the *EYA4* transcript. The subtle differences in clinical presentations for rarer, more severe cases as opposed to the more common hearing

phenotypes could be a result of variants in different functional regions of the gene resulting in a range of phenotypes.

CDH23 is another example of a gene that causes varied phenotypes based the location of multiple mutations, ~116 sites in *CDH23* contain variants that cause auditory phenotypes. For example, some *CDH23* USH1D patients present with both visual and auditory phenotypes while others have auditory phenotypes only. In addition, variation between phenotypes may be caused by variants that are involved in gene-gene interactions.

With respect to using current knowledge of these genes to better understand ARHI, expression has been tested for six of these genes in the peripheral auditory system. The auditory structures in which these genes are expressed are therefore candidates as sites for biological mechanisms implicated in *HDiff* and *HAid*. Six have known roles in the cochlea, mainly in hair cell (HC) stereocilia function and maintenance: *SPTBN1,* HC stereocilia[296]; *TRIOBP,* HC stereocilia[53]; *SYNJ2,* HCs[318]; *LMX1A,* endolymphatic duct and basal HCs[339]; *CDH23,* HC stereocilia[55]; *CTBP2,* HCs[341]. *TUB* has a role in in the function of neuronal cell synapses[337] while *EYA4* is known to have a role at the tympanic membrane[336]. However, while six of the eight genes have a role in the cochlea, this does not necessarily indicate that the cochlea is the only site of pathology, or that expression is not altered under different conditions.

Lastly, a number of these ten genes have been linked to other conditions that are associated with ARHI. *SYNJ2* for example, has been linked to Alzheimer's[319,342] and depression[343,344], and is known to have a functional role in neuronal cells. With this knowledge, it could be hypothesised that *SYNJ2* function may be in higher auditory structures and that their degeneration leads to hair cell dysregulation and death. The hypothesis is further supported by the observation that the *Synj1* knock-out mouse display neurological defects and die shortly after birth. SYNJ1 and SYNJ2 are two key proteins in the Synptojanin protein family.

Although a *Synj2* knock-out mouse has not yet been made, the hypothesis that a neurological defect is the underlying cause of HL was tested, in the *mozart* (Synj2 mutant mouse as described on p.154). the study performed nerve conductance tests on *mozart* mice yet found no abnormal neurological behaviour at age 12 weeks, by which time *mozart* mice were severely deaf[318].This indicates that *mozart* mice do not display a neurological phenotype primary to hair cell degeneration and loss and that the underlying cause of HL in *mozart* may therefore be the inner ear defects. Phenotypes of the aged heterozygous mutant were not assessed beyond 12 weeks however, and authors note the multiple functional regions

in the gene that may have diverse functions[318]. Determining the role of associated genes in other conditions can improve knowledge of the gene role in hearing and the biological relationship between the two or between multiple traits.

The associations observed at these loci are not 'novel' in respect to a link with hearing or auditory function. In addition, based on the data presented in this chapter, all of these targets have been previously characterised and therefore none was selected for use in immunohistochemistry analysis at this stage. For most of the genes however, this is the first link to common hearing impairment in a human population and so these are important findings for human ARHI genetic risk. The findings also support the use of mice to study human ARHI genetic risk, challenges current understanding of the function of these genes, and could be interpreted as a validation of the association analysis and reliability of phenotypes.

5.2.3 Loci associated in *HDiff* and *HAid* genetic association analysis that contain genes that are novel associations with hearing function or hearing conditions

In this section the 34 remaining loci are evaluated. These are the loci that have not previously been linked to auditory function or pathologies and I selected three candidates for protein localisation analysis in adult mouse cochlear tissue. As noted at the beginning of this chapter, i*n vitro* and *in vivo* investigation of gene candidates requires multiple resources and can be a time-consuming process. Therefore, prioritisation of gene candidates is an essential step and ought to be thorough, with bias omitted where possible. For each locus, data gathered in section 5.2.1, (where the genomic context of lead SNPs and thus the confidence of gene candidate selection was reported), is extended using further criteria set out in Figure 5.2. The results are summarised in Table 5.5.

Here the statistical strength and validity of the associations is summarised. Each candidate is assessed based on; the lead SNP p-value, whether the locus is significantly associated in both *HDiff* and *HAid* and whether the association is significant in the replication meta-analysis, Table 5.5, columns 3, 6 & 7. Each candidate is also assessed based on the beta coefficient; the relative effect of allele copy number on the phenotype variation, Table 5.5 column 4. As the association analysis uses qualitative data, the beta value is not accurate for assessing the exact 'effect' of the variant on phenotype variation but can be used for making comparisons between SNPs within this study.

In Table 5.5, column 12, it is noted whether any of the gene candidates had putative links to hearing pathologies or the auditory system. Data from the gEAR portal (https://umgear.org/) was also investigated in relation to the primary gene candidates. This data includes gene expression levels in hair cells and supporting cells from P0 mouse cochlear samples in recent RNA-seq screens[345] (- Methods 2.4.3). These findings can be interpreted as a form of 'validation' of firstly the association at the locus, and secondly, confidence in the selection of the true functional candidate.

Table 5.5 also includes information regarding the availability of mutant mouse models and whether these display a hearing phenotype (in cases where such a test had been performed, column 10). This data is included as, if the relevant mutant mouse displays a hearing phenotype, this provides evidence of the gene causing a hearing pathology and that this function is conserved across humans and mice. If the strain is available, the pathophysiology can be readily studied. It is also noted in Table 5.5 whether antibodies are commercially available against the primary candidate protein product, and if it has been reported in a peer-reviewed publication, Table 5.5, column 11.

In order to summarise and rank information for each gene candidate, Table 5.5 is colour coded; dark blue signifies strong evidence to support gene prioritisation, pale blue signifies some evidence to support gene prioritisation and grey signifies relatively low evidence for gene prioritisation. To minimise the duplication of information within the table, if multiple SNPs were assigned the same primary candidate gene in section 5.2.1 (the three SNPs at the *ARHGEF28* locus and the two SNPs at the *NID2* locus) these SNPs are grouped together in the table within this section.

The data in Table 5.5 is interpreted and discussed in the text that follows. Evaluation of the data is subset into four sections for ease of assessment. Gene candidates are subset into groups based on parameters within the table. Some of the candidates could be assigned to more than one group such as high statistical evidence from association and high expression in inner ear sequencing screens. In these instances, the candidate is listed in the first section to which it can be assigned.

Lead SNP	Genomic position	P-Value	Beta (SE)	Primary candidate	Pheno.	Rep.	Mouse	gEAR, NSC/HC (RPKM)	Mouse pheno.	Anti.	Link to hearing
rs36062310	Exon	1.90E-22	-0.0315 (0.003)	*KLHDC7B*	*HDiff*	✓	✓	20 / 8402.9	-	✓-	-
rs6453022	Exon	1.70E-21	-0.0126 (0.001)	*ARHGEF28*	*HDiff*	-	✓	No data	-	✓-	-
rs6890164	Intergenic	3.30E-19	0.0119 (0.001)		*HDiff*	-					
rs4597943	Intron	2.09E-11	-0.0042 (0.001)		*HAid*	✓✓					
rs759016271	Intron	6.10E-21	-0.0127 (0.001)	*ZNF318*	*HDiff*	✓✓	✓	No data	-	✓-	-
rs7951935	Intergenic	7.80E-17	-0.0114 (0.001)	*TYR*	*HDiff*	✓	✓	19420.4 / 331	✓	✓✓[346]	Possible protective role via melanin[347]
rs35186928	Intergenic	1.70E-15	-0.0109 (0.001)	*HLA-DQA1*	*HDiff*	✓	✓	No data	-	✓-	-
rs1566129	Intron	1.40E-11	0.0091 (0.001)	*NID2*	*HDiff*	✓✓	✓	329277 / 24934.7	-	✓✓[348]	-
rs1566129	Intron	2.50E-09	0.0037 (0.001)		*HAid*	-					
rs35414371	Intergenic	1.60E-11	-0.0131 (0.002)	*CLRN2*	*HDiff*	-	✓	20 / 424.2	✓-*	✓-	Function at HC stereocilia[349]
3:182069497_TA_T	Intergenic	4.10E-11	-0.0118 (0.002)	*ATP11B*	*HDiff*	-	✓	2649.9 / 5705.9	-	✓-	-
rs12225399	Intron	8.60E-11	-0.009 (0.001)	*PHLDB1*	*HDiff*	✓	✓	31373.4 / 9133.8	-	✓-	-
rs62033400	Intron	2.90E-10	0.0085 (0.001)	*FTO*	*HDiff*	✓	✓	11271.5 / 26941.5	-	✓✓[350]	-
rs13277721	Intron	3.30E-10	-0.0083 (0.001)	*AGO2*	*HDiff*	-	✓	No data	-	✓✓[351]	-
rs62188635	Intergenic	4.70E-10	0.0083 (0.001)	*KLF7*	*HDiff*	✓	✓	809.5 / 327.4	-	✓✓[352]	-
rs4947828	Intron	1.00E-09	-0.0096 (0.002)	*GRB10*	*HDiff*	-	✓	1214.1 / 482.6	-	✓-	-
rs34442808	Intron	1.30E-09	-0.008 (0.001)	*MCTP1*	*HDiff*	-	✓	81 / 1475.3	✓	✓-	Contains a cis-regulatory region for NR2F1 function in mouse inner ear development[347]

Lead SNP	Genomic position	P-Value	Beta (SE)	Primary candidate	Pheno.	Rep.	Mouse	gEAR, NSC/HC (RPKM)	Mouse pheno.	Anti.	Link to hearing
rs835267	Intron	1.60E-09	0.008 (0.001)	*EXOC6*	*HDiff*	-	✓	256.4 / 2402.8	-	✓-	-
rs4948502	Intron	1.70E-09	0.0081 (0.001)	*ARID5B*	*HDiff*	-	✓	13534.1 / 2250.6	-	✓✓[353]	-
rs10824108	Intron	3.00E-09	-0.0079 (0.001)	*ADK*	*HDiff*	✓	✓	3407.9 / 2317	-	✓✓[354]	-
rs12027345	Intergenic	3.60E-09	0.0079 (0.001)	*MAST2*	*HDiff*	✓	✓	20 / 20	-	✓✓[355]	-
rs217289	Intron	4.90E-09	-0.0078 (0.001)	*SNAP91*	*HDiff*	✓	✓	723.9 / 4711.6	-	✓-	-
rs13093972	Intergenic	5.50E-09	-0.0078 (0.001)	*ZBTB20*	*HDiff*	-	✓	20 / 20	-	✓✓[356]	-
rs62015206	Intergenic	7.70E-09	-0.0078 (0.001)	*MAPK6*	*HDiff*	-	✓	20 / 33.3	-	✓✓[357]	-
rs10475169	Intergenic	9.30E-09	-0.0117 (0.002)	*IRX2*	*HDiff*	-	✓	1931.5 / 10284.9	-	✓-	-
rs17671352	Intron	1.00E-08	0.0078 (0.001)	*ACADVL*	*HDiff*	✓	✓	4967.8 / 5301.5	-	✓-	-
rs12938775	Intron	1.60E-08	0.0075 (0.001)	*PAFAH1B1*	*HDiff*	-	✓	192.4 / 376.5	-	✓✓[358]	Evidence of a role in HC planar polarity[358]
rs76837345	Intron	1.90E-08	-0.0146 (0.003)	*CHMP4C*	*HDiff*	-	✓	82.1 / 297.4	-	✓	-
rs9366417	Intron	2.10E-08	0.0085 (0.002)	*SOX4*	*HDiff*	-	✓	12907.2 / 10772.7	-	✓✓[359]	SoxC gene family with role in HC production[360]
rs3890736	Intergenic	2.20E-08	-0.0077 (0.001)	*GFRA2*	*HDiff*	-	✓	24971.7 / 1987.8	-	✓✓[361]	Expression in spiral neuron cell bodies[362]
rs7823971	Intergenic	2.70E-08	-0.0043 (0.001)	*RP11-1102P16.1*	*HAid*	-	-	No data	-	✓-	-
rs9691831	Intergenic	3.10E-08	-0.0074 (0.001)	*TMEM213*	*HDiff*	-	✓	20 / 20	-	✓-	-
rs141403654	Intron	3.50E-08	-0.0313 (0.006)	*AGBL2*	*HDiff*	✓	✓	154.6 / 718.3	-	✓-	-
rs4611552	Intergenic	3.60E-08	-0.0089 (0.002)	*CCDC68*	*HDiff*	-	✓	23 / 541	-	✓-	-
rs12552	Exon	4.80E-08	0.0073 (0.001)	*OLFM4*	*HDiff*	-	✓	67.3 / 82.5	-	✓✓[363]	-
rs10927035	Intron	4.90E-08	-0.0075 (0.001)	*AKT3*	*HDiff*	-	✓	3095.1 / 11550	-	✓-	-

Lead SNP	Genomic position	P-Value	Beta (SE)	Primary candidate	Pheno.	Rep.	Mouse	gEAR, NSC/HC (RPKM)	Mouse pheno.	Anti.	Link to hearing

Table 5.5. Summary table of data for primary gene candidates that are novel to hearing pathologies.

Lead SNP; lead SNP at each region of association as defined by joint and conditional in chapter 4, Genomic position; genomic position of lead SNP, P-Value; BOLT-LMM infinitesimal mixed model association test p-value of the lead SNP in GWAS presented in chapter 4, Beta (SE); effect size from BOLT-LMM approximation to infinitesimal mixed model and (SE) standard error of the beta effect size, Primary candidate; gene in closest proximity to lead SNP, Pheno.; phenotype used for GWAS, Rep.; ✓ denotes a nominal p-value in the meta-analysis replication and ✓✓ denotes a Bonferroni-corrected p-value in the meta-analysis replication, Mouse model; ✓ denotes that there is a mutant mouse model for this gene inc. archived strains, gEAR NSC/HC, RPKM; Results from RNAi screen of P0 mouse expression profile for non-sensory cells (NSC), Hair cells (HCs) Reads Per Kilobase of transcript, per Million mapped reads (RPKB), Mouse hearing phenotype; ✓ denotes that the mutant mouse model displays a hearing phenotype, Anti; ✓ denotes that there is a commercially available antibody for the protein of interest while ✓✓ denotes that the antibody has been used in a peer-reviewed publication, Link to hearing; details of any putative link to hearing in the current literature. Dark blue signifies strong evidence to support gene prioritisation, pale blue signifies some evidence to support gene prioritisation and grey signifies relatively low evidence for gene prioritisation. * denotes a finding that was not published when the work was conducted.

5.2.3.1 Gene candidates with strong statistical evidence based on association analysis

Three primary candidates, *ARHGEF28*, *ZNF318* and *NID2,* have strong statistical support for prioritisation, based on the association statistics from this study. Gene transcripts of all three candidates contain lead SNPs P<1.5E-11 in the discovery analysis, and all three replicate at a Bonferroni-corrected significance threshold in the replication meta-analysis. In addition, *ARHGEF28* and *NID2* are primary candidates that contain SNPs significantly associated (P<5E-08) with in both of the two phenotypes *HDiff* and *HAid.*

In addition to this strong statistical evidence, gene family members of *NID2* and *ARHGEF28* have been linked to hearing function or have been shown to be expressed in auditory structures. *ARHGEF6* is in the same guanine nucleotide exchange factor family as *ARHGEF28* and was recently studied in the context of its role in sensorineural hearing loss. *ARHGEF6* was hypothesised to have a role in sensorineural hearing loss as is also known as the gene which causes x-linked intellectual disability (as *XLID46*), a disorder where sensorineural hearing loss can present as a symptom[364]. A recent study identified expression of Arhgef6 in the mouse cochlea and used a knock-down model to test for a functional role. The knock-down mice displayed a significant threshold difference to wild-type siblings at P90 and by P60, exhibited abnormal morphology of IHC hair cell bundles and exhibited abnormal morphology of OHC bundles[365]. Being of the same gene family, it is possible that *ARHGEF28* has a similar role as, or site of action within the cochlea to, *ARHGEF6*.

Expression of *NID1* has been observed in cochlear structures[366] and, as NID1 and NID2 are both components of the basement membrane structure, it could be hypothesised that NID2 would also be present and have a function in the structure. The exact role of NID1 in the cochlea has not been determined, but expression was observed lining blood vessels, in the Reissner's membrane and the osseous spiral lamina. As the two proteins are expressed as components of the same basement membrane structures, it is likely that *NID2* could also be expressed in the cochlea.

5.2.3.2 Gene candidates with putative previous links to hearing function or hearing conditions

Six of the candidates; *TYR, CLRN2, MCTP1, PAFAH1B1, SOX4* and *GFRA2,* either have putative links to hearing reported in the literature that do not meet the criteria used in section 5.2.1 or links that have subsequently been published. None of the six genes has been directly linked to auditory pathology in humans, though three have been linked to mouse hearing

phenotypes; *Tyr*, *Clrn2* and *Mctp1*. Therefore, these is an interesting subset of potential candidates for further investigation. Below the individual merits and limitations of each candidate is discussed.

Firstly, the lead SNP in close proximity to *TYR* (rs7951935) has a low p-value p=7.80E-17, and nominal replication in the replication meta-analysis. The lead SNP at this locus however is not within the *TYR* transcript, and although *TYR* is the nearest protein-coding gene transcript, there are multiple genome-wide associated SNPs that are in high LD with the lead SNP that reside in the *NOX4* transcript, along with non-significant SNPs that are in moderate LD, in the *GRM5* gene transcript. Figure 5.4. At this locus fine mapping is required to gain more confidence in gene candidate selection for functional analysis.

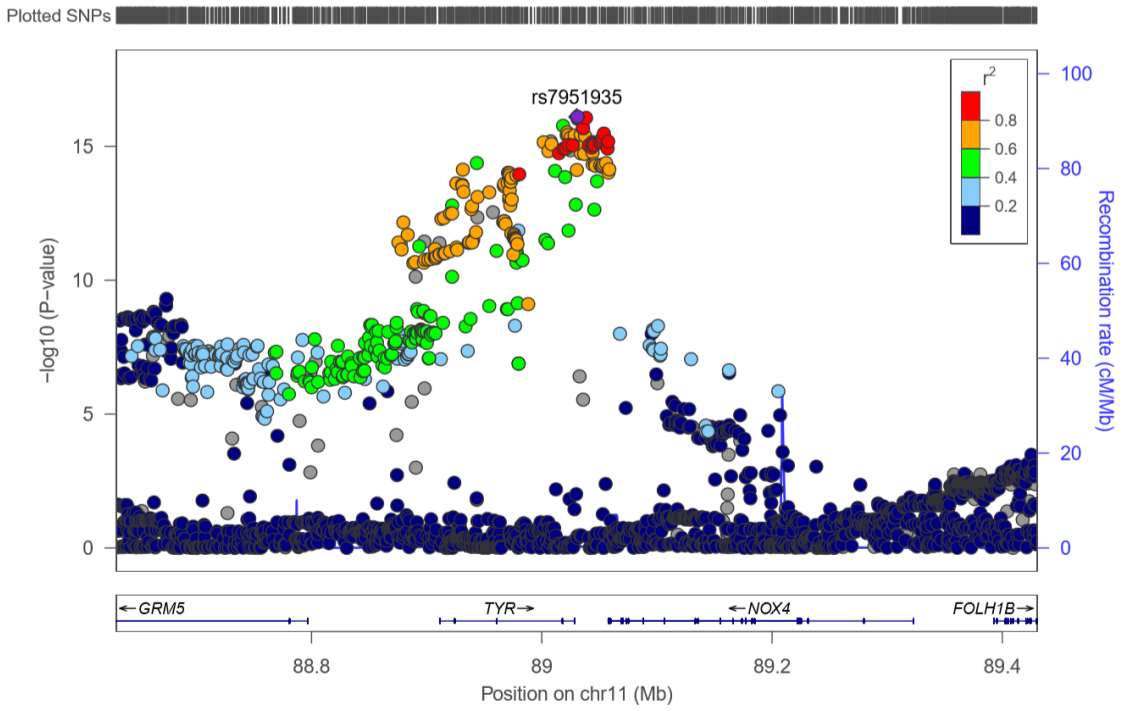

Figure 5.4. Locus plot displaying the genome region flanking the lead SNP rs7951935, in proximity to the *TYR*, *NOX4* and *GRM5* transcripts.

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey.

Speculated links between tyrosinase (*TYR)* and hearing loss have been made, relating to the role of tyrosinase in the conversion of tyrosine to melanin. The production of melanin is thought to be a protective factor against hearing loss, specifically noise-induced hearing loss. To explore these observations, albino mice were studied in comparison to wild-type and to tyr-transgenic mice, where it was observed that melanin precursors had a protective effect against sensorineural hearing loss following noise exposure damage[99].

Secondly, a lead SNP in close proximity to *CLRN2* has an association p-value of 1.60E-11 and is located 1956bp downstream of the *CLRN2* gene transcript. Within the same LD block is the *QDPR* gene transcript. While *QDPR* has no previous link to hearing loss, *CLRN2,* which codes for clarin-2, is a strong candidate for two reasons. Firstly, *CLRN2* is a paralogue of *CLRN1*, a gene which, when mutations are present, is known to cause autosomal recessive Usher syndrome Type-3 (sensorineural hearing loss is present from birth)[367,368]. Based on this, it could therefore be hypothesised that CLRN2 may also have a function in auditory structures.

Furthermore, a recent study using *Clrn2*$^{clarinet/clarinet}$ mice generated via a large-scale forward genetic screen, demonstrated that *Clrn2* is required for normal hair bundle integrity and functioning[349]; the mutant mice displayed a loss of mechano-electrical transduction followed by a gradual loss of HC stereocilia. Figure 5.5 depicts the loss of IHC and OHC stereocilia in p28 mice. Genetic association analysis that I performed using the *HDiff* phenotype (described in chapter 3 and 4), on 484 SNPs within 100kb of the *CLRN2* transcript, was also described in the study. The association in the UKBB cohort compliments the finding in the *Clrn2*$^{clarinet/clarinet}$ mice of CLRN2 function and demonstrates its relevance to human auditory function. However, protein localisation of CLRN2 in the cochlea has not yet been established via immunohistochemistry, so this would confirm the location of expression and support the evidence of this specialised function, and test whether the association here likely relates to the function of *CLRN2* or *QDPR*.

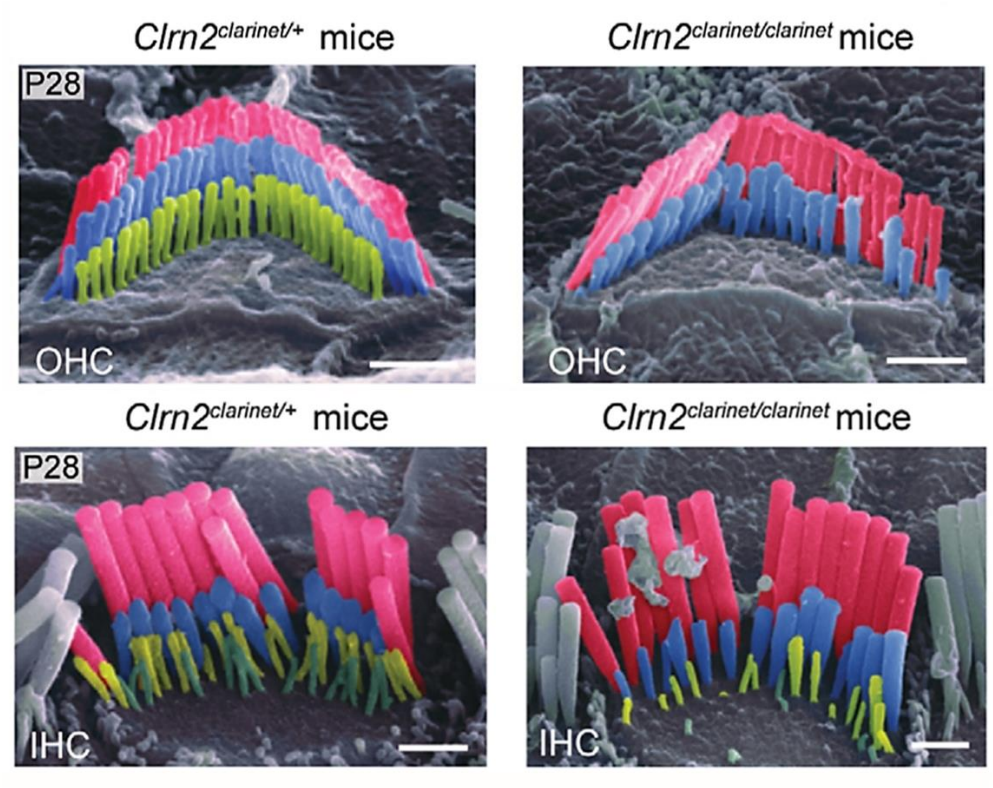

Figure 5.5 Loss of the OHC and IHC hair bundle transducing stereocilia in *Clrn2*$^{clarinet/clarinet}$ mice.

Pseudo-coloured scanning electron micrographs of individual OHC (top panels) and IHC (bottom panels) hair bundles from *clarinet* mice at P28. *Clrn2*$^{clarinet/clarinet}$ mutants have only two rows of OHC stereocilia, and the middle row of stereocilia is less uniform in height compared to controls. The middle and short rows of stereocilia in *Clrn2*$^{clarinet/clarinet}$ IHC bundles appear fewer in number, and heterogeneous in height. Scale bars, 2µm. Figure and legend reproduced and adapted with permission, L. Dunbar et al.[369].

Thirdly, the lead SNP that resides in *MCTP1* is also highly associated with the *HDiff* p=1.30E-09, but the association is not replicated in the meta-analysis. The lead SNP here resides in an intron region of *MCTP1*, indicating that this gene may be the functional target of the causal variant. However, there are also multiple SNPs at genome-wide significance that are in high LD with the lead SNP, that reside in the *ANKRD23* gene transcript, downstream of *MCTP1*. *MCTP1* does however have a putative link to auditory function as it is understood to contain a cis regulatory region that controls the activity of *NR2F1*, a gene that has a role in the development of the sensory hair cells and cochlear epithelium[370].

MCTP1 is therefore an example of a gene highly associated with a phenotype of interest but does not have a direct functional role in the auditory system. Instead, via a regulatory interaction, it affects the behaviour of a gene that is directly involved in cochlear development[370]. Therefore, protein localisation of MCTP1 in the cochlea would not be relevant to perform here. Secondly, fine mapping of this region is required to affirm the

assumption that *MCTP1* is the target of the observed association at this locus rather than *ANKRD23*.

The remaining three genes, *GFRA2*, *PAFAH1B1* and *SOX4* have been linked to auditory function by expression and, or function within critical auditory structures. By comparison to the other lead SNPs studied here, *GFRA2* has a relatively lower significance level of association (p=2.20E-08) and is not replicated at nominal significance. Expression of *GFRA2* has been detected in adult human cochlear samples at spiral neuron cell bodies[99], suggestive of a potential function in spiral ganglion neurons. This is evidence that *GFRA2* is a strong candidate to pursue for determining a pathology, but as the location in the cochlea has already been established it is not prioritised for immunohistochemistry at this stage.

Similar to *GFRA2*, the p-value for the lead SNP in the region of *SOX4* is relatively low (p=2.10E-08) by comparison to the other lead SNPs in this analysis. *SOX4* is a member of the SOXC transcription factor (TF) family which has been identified as potential targets for regeneration of supporting and hair cells in the mammalian inner ear[360]. This is a result of the finding that SOXC TFs (Sox4and Sox11*)* are strongly downregulated from E17.5 to P9 in the mouse inner ear. Further to this, proliferation of progenitor cells expressing Sox4 and Sox4 was observed *in vivo* in E16.5 mice[360]. In the same study, *Sox4* expression was observed at hair cells via immunohistochemistry. Therefore, *SOX4* is not a prioritised candidate for immunohistochemistry here, though expression could be present at additional structures in the cochlea in adult mice.

In addition, the lead SNP at the *SOX4* locus in this analysis may not actually represent a pathogenic function in the *SOX4* gene. While *SOX4* is the nearest protein-coding gene, the lead SNP is upstream of the *SOX4* transcript and, as mentioned in section 5.2.1, is located in the RNA gene *CASC15*, in a different LD block to the *SOX4* transcript (Figure 5.6). A fine mapping approach could be used to explore whether *SOX4* is possibly the functional target of the pathogenic variant or the regulatory element *CASC15*.

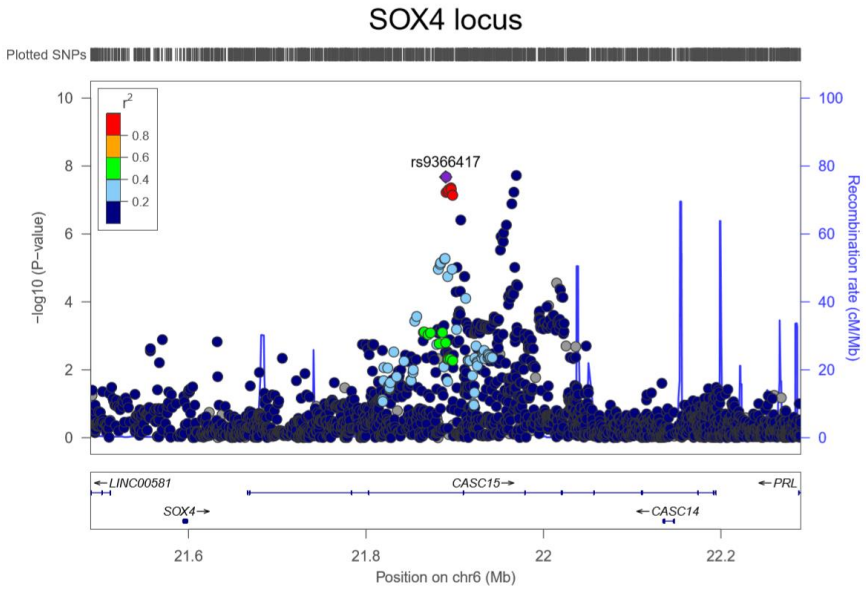

Figure 5.6. Locus plot displaying the SOX4 protein-coding transcript and the CASC15 RNA-coding transcript.

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey.

Lastly, the lead SNP that resides in an intron of *PAFAH1B1* has an association p-value of 1.60E-08, and while it does not reach nominal significance in the replication analysis, no other protein coding genes at the locus contain genome-wide associated SNPs in high LD with the lead SNP, giving confidence to the assertion that *PAFAH1B1* is the functional target. *PAFAH1B1* has been linked to hearing function firstly based on a hypothesis that it has a role in determining hair cell planar polarity. HC planar polarity is crucial to sensory cell function, as the precise organisation of stereocilia on the apical end of sensory cells affects their mechanostransduction.

PAFAH1B1 (Lis1) is a microtubule regulator and thus was selected as a candidate regulatory gene for HC planar polarity. Based on the hypothesis, HC formation and maintenance has been studied in conditional *Lis1* mutant mice. At the embryonic stage, the mutant mice display defective organisation of cells at the organ of Corti and of HC planar polarity. Maintenance of HC planar polarity was also affected in postnatal cochlear tissue in the mutants via the regulation of cytoplasmic dynein and organization of microtubules[358]. In this study too, immunohistochemistry was used to visualise structures within the cochlea and thus this candidate is not prioritised for immunohistochemistry here.

5.2.3.3 Gene candidates that are a novel association with hearing function or hearing conditions, yet display relatively high mRNA expression in the inner ear

Of the remaining candidates, 12 have relatively high levels (>1000 RPKM) of gene expression in either NCS or HCs in P0 mouse inner ears, (Table 5.5, column 9). The data were collected from an RNA-seq screen of Atoh-1 expressing HCs. Within this group of gene candidates to have relatively high expression levels in this screen is *KLHDC7B.* The lead SNP at the *KLHDC7B* locus is rs36062310 and has the smallest p-value of all lead SNPs in the two association analyses P=1.90E-22. The lead SNP resides in an exon region of *KLHDC7B,* Figure 5.7*.*

In addition to the relatively high expression in the RNAi screen, the protein products of *Klhdc7b* and *Adk* has been observed in a hair cell proteome screen using P4-P7 Pou4f3/eGFP-transgenic mice. KLHDC7B is one of 934 proteins identified in the HC sample only (GFP + sample)[371]. Adk expression was also observed in the GFP + sample and was one in a group of 38 HC-only protein isoforms; a number of the genes studied have multiple isoforms which displayed differential expression in cochlear cell types examined. In contrast to this however, expression of *Klhdc7b* and *Adk* was not particularly high in the RNA-seq screen of Atoh1-expressing HCs in P0 mice. Table 5.5 displays the expression level of these transcripts in HCs in this screen as 2317 RPKM (*Adk*) and 8402.9 RPKM (*Klhdc7b*).

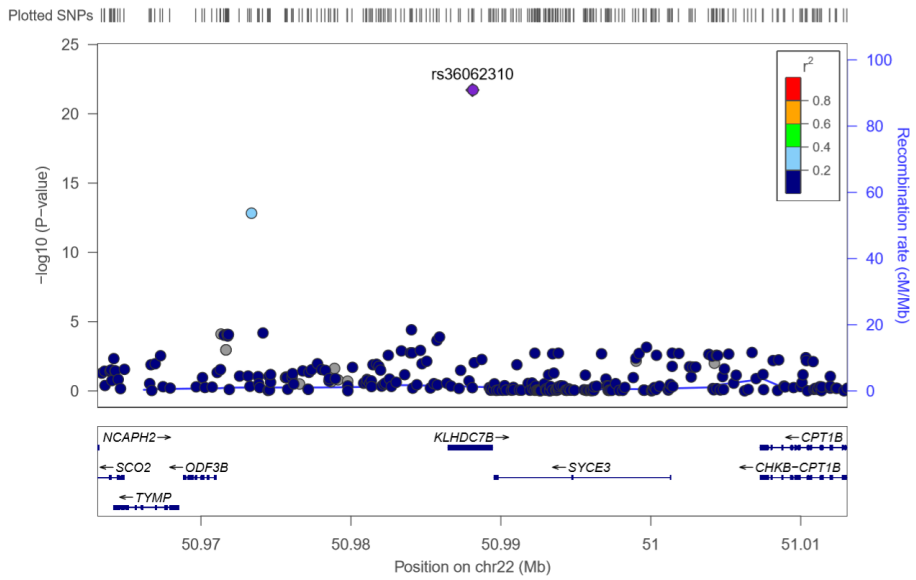

Figure 5.7. Locus plot of region containing the lead SNP rs36062310 p=1.90E-22 within an exon of *KLHDC7B*.

The plot displays the region flanking the lead SNP (purple) by +/- 25kb. The colouring of remaining SNPs represents the correlation (r2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey.

Within this group of 12 candidates, five have very high relative levels of expression in the RNA-seq screen[261]; *Phldb1, Fto, Arid5b, Irx2* and *Gfra2.* Lead SNPs at *PHLDB1, FTO, ARID5B* loci lie within intron regions of respective genes, and no further SNPs at the loci reached genome-wide significance and are in high LD with the lead SNP. For these candidates, functional analysis could be justified, but will not be performed at this stage due to the stronger evidence of association and function for other candidates discussed previously. For *IRX2* and *GFRA2* however, there is less confidence in the selection of the primary candidate from the lead SNPs. The lead SNPs at the *IRX2* and *GFRA2* loci are >100kbp and >10kbp away from the gene transcripts respectively. Fine mapping would be required to confirm whether these are the most likely functional target of the associations at these loci.

5.2.3.4 Gene candidates that are a novel association with hearing function or hearing conditions that display a relatively low mRNA expression in the inner ear

Twelve candidates remain that are not in the top three loci with the strongest statistical evidence of association, have no other putative links to hearing function and have <1000 RPKM in NSCs and/or HCs in RNA-seq screen[261] (or were not included in this screen). For these candidates, further analysis is necessary before allocating resources to follow-up any of the targets *in vivo*. For example, this group contains the primary candidate RP11-1102P16.1, an uncharacterised protein coding gene.

In addition, a number of these lead SNPs are some distance from the nearest gene transcript. An example is depicted in Figure 5.8, which displays the region flanking rs13093972, a SNP in proximity to the *ZBTB20* gene transcript. This could be evidence that the functional variant has a regulatory role either upstream or downstream of the coding gene target; again a likely hypothesis in ARHI, as variants within the population collectively have an effect based on relatively small individual effects, such as observed when gene regulation is subtly altered. So, while there is not sufficiently strong evidence that these candidates and lead SNPs are suitable to prioritise for protein expression in the mouse cochlea at this time, these are important findings and ought to be evaluated again and may be suitable for follow-up using alternative methods.

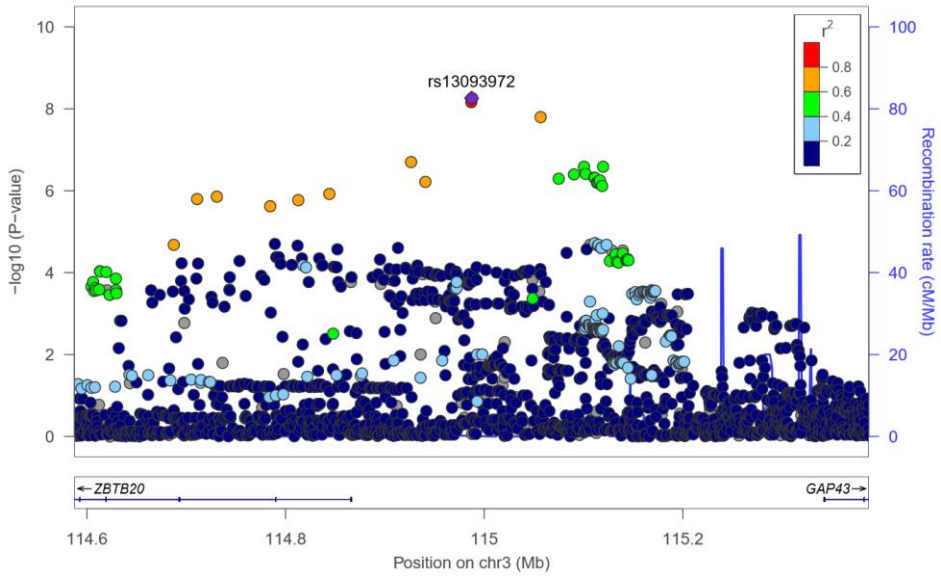

Figure 5.8. Locus plot for the region containing the lead SNP rs13093972 p=5.50E-09 in proximity to ZBTB20.

The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays the region flanking the lead SNP (purple) by +/- 400kb

Based on the distance of lead SNPs from some of these primary candidates, the reported low expression in the P0 mouse inner ear RNA-seq screen and an absence of previous putative links with these candidates, these may not be the functional targets of the causal variant. Conversely, as these are novel to hearing loss, these could be the most progressive findings. Furthermore, the lack of expression in the RNA-seq screen could indicate that the genes are not expressed during development and therefore have a maintenance function, as expected in a progressive and heterogeneous trait such as ARHI.

5.2.4 Immunohistochemistry

In the previous sections, gene candidates were evaluated with the aim of selecting a subset for protein localisation in adult cochlear tissue. In this final section, the selection of candidates and subsequent immunohistochemistry results are presented. The corresponding methods for protein localisation analysis are described in - Methods 2.5. The three loci selected contain the genes (1) *NID2*, (2) *ARHGEF28* and (3) *CLRN2*, and are annotated on the Manhattan plots in Figure 5.9.

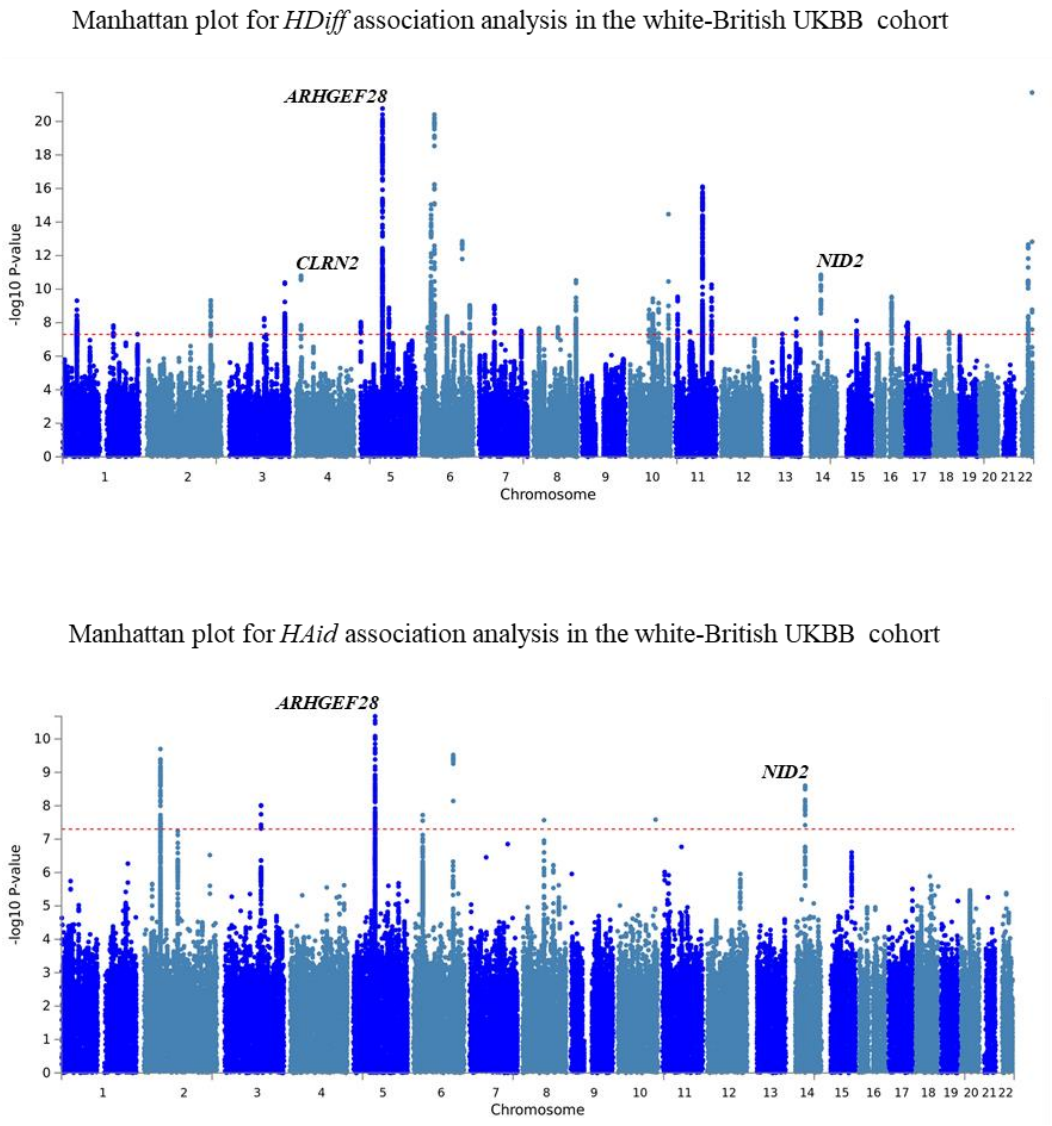

Figure 5.9. Two Manhattan plots displaying GWAS results for *HDiff* (top) and *HAid* (bottom).

The graphs are annotated with the position of the three genes that code for ARHGEF28, CLRN2 and NID2, which were selected for immunohistochemistry. The red dotted line marks genome-wide significance p<5E-08. The loci containing *NID2* and *ARHGEF28* contained SNPS that reached genome-wide significance in both *HDiff* and HAid. SNPs at the *CLRN2* locus reached genome-wide significance in *HDiff* only.

Expression of all three proteins is tested in n=3 adult mouse cochlear tissue (p28-p30), following the protocol set out in Chapter 2, section 2.5 'Protein localisation analysis'. In

addition to antibodies against the proteins of interest, phalloidin and DAPI are applied to samples to stain F-actin and AT regions of dsDNA respectively. F-actin is present in hair cell stereocilia and is therefore here used as a marker to locate the cochlear sensory hair cells. DAPI stains AT regions of dsDNA, mainly present in the nuclei, and thus can be used to visualise the structure of the sample from the composition of individual cells.

5.2.4.1 Nidogen-2 (NID2)

Nidogen-2, coded by *NID2* was selected for protein localisation analysis. The same lead SNP rs1566129 reached genome-wide significance in both *HDiff* and *HAid* work and is located in intron 5 of *NID2*. The associated region can be observed in Figure 5.10 (*HDiff*) and Figure 5.11 (*HAid*) with the lead SNP coloured purple. Within the loci, *NID2* is the only gene transcript that contains SNPs that are within high LD (r^2>0.6) at this locus.

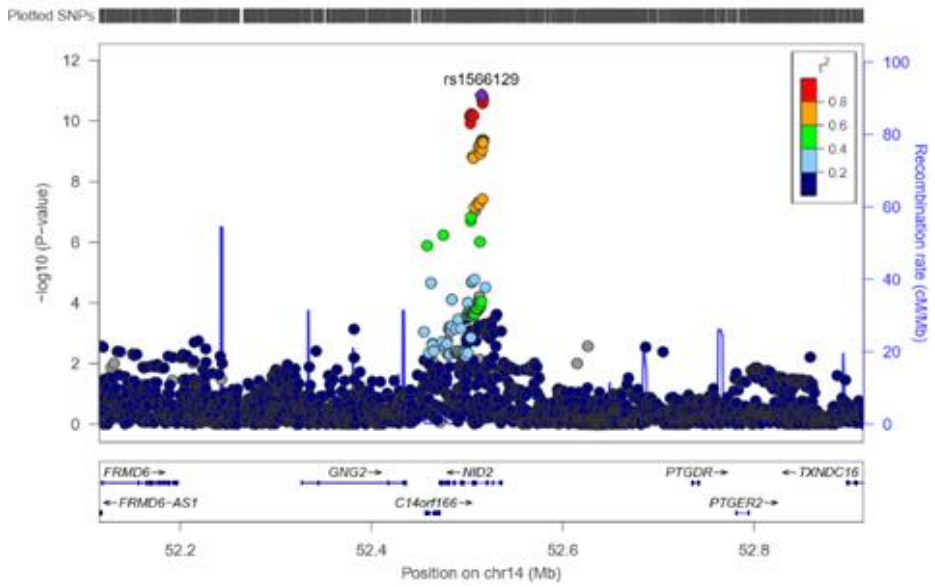

Figure 5.10 Locus plot of *HDiff* lead SNP and the NID2 gene transcript.

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 100kb +/- regions flanking the position of the lead SNP.

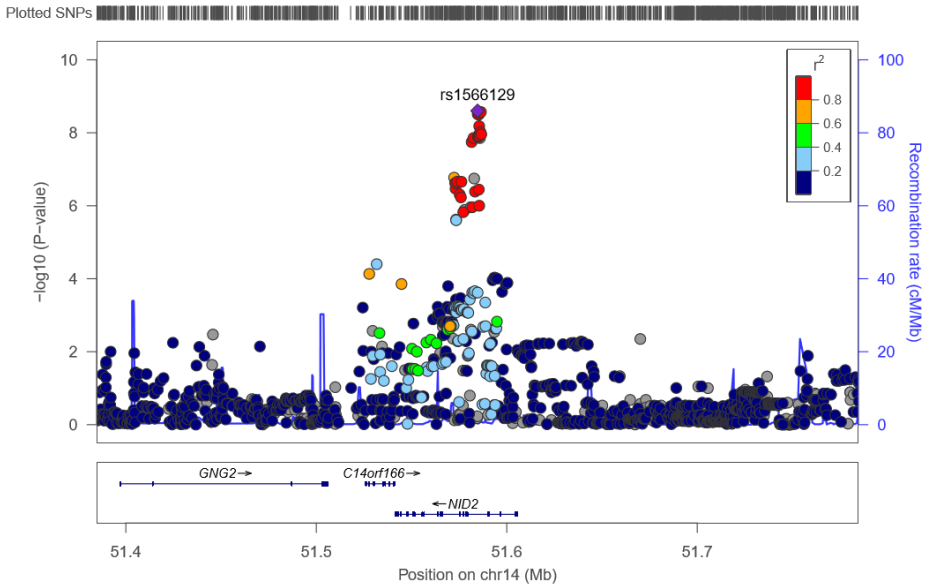

Figure 5.11 Locus plot of *HAid* lead SNP and the *NID2* gene transcript

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 50kb +/- regions flanking the position of the lead SNP.

Nidogen-2 is part of the nidogen glycoprotein family and is a component of the basement membrane complex, along with nidogen-1, the first isoform of the protein to be discovered. In the basement membrane complex, nidogen-2 binds to laminin-1 and to collagen I and IV[372]. Nidogen-2 is enriched in endothelial basement membranes and has a role in the formation and homeostasis of blood vessels. Work in *Nid2*-mutant mice has revealed that the two proteins may not have a structural role in the basement membrane as its formation was not affected[373]. Findings from another study that focused on *Caenorhabditis elegans* which possess one nidogen-coding gene, suggest that nidogens could have further roles, in addition to being part of the basement membrane complex. Nidogen mutants did not exhibit abnormal basement membranes but display affected axonal patterning such that the dorsal sub-lateral nerves are mispositioned to the dorsal midline[374].

The role of nidogens in auditory function has not previously been examined and there have been no previous associations with nidogens and auditory pathologies. The only link that has been previously published has been the identification of nidogen-1 expression in the supporting cells of the organ of Corti, the Reissner's membrane and the osseous spiral lamina [366]. There are commercially available antibodies that have been successfully used in publications for protein localisation of nidogen-2 in multiple tissue types[375].

Protein localisation of NID2

Here, anti-NID2 staining was observed in the lining of blood vessels, visible in Figure 5.12 (c) bottom arrow, and (b) along the structure of the stria vascularis). This is evidence of antibody specificity in these samples, as nidogen-2 has previously been observed in blood vessel basement membranes in several tissue types[372]. Anti-NID2 staining is most prominent in a restricted region of the epithelial lining between the inner spiral sulcus and spiral limbus (Figure 5.12), with strong staining at the base of a cell in proximity to the inner hair cell (Figure 5.12 (c) top arrow and Figure 5.13). In the 40x images (Figure 5.13), anti-NID2 staining is again observed in this region, appearing to be around the base of one cell at each plane of the z-stack. Anti-NID2 staining is not observed in any of the cell bodies of the sensory cells.

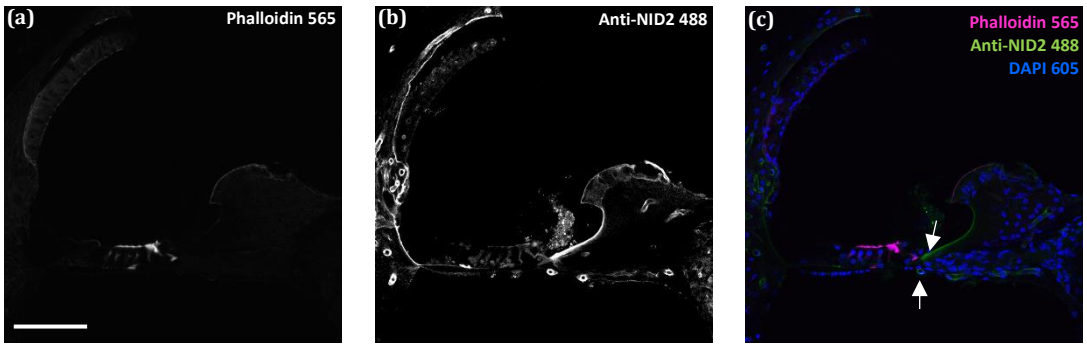

Figure 5.12. Protein localisation of Anti-NID2 at the organ of Corti, 20x objective.

Colour labelling in the composite image (c): Blue, DAPI; Magenta, phalloidin; Green, Anti-NID2. The white arrow directing downwards in (c) is to highlight prominent staining (anti-NID2, green) at a region of the epithelial lining between the inner spiral sulcus and spiral limbus. The white arrow pointing upwards is to highlight the blood vessel lining; the circular structure stained by anti-NID2 (green). The scale bar in (a) represents 100µm and is consistent for all three images in this figure.

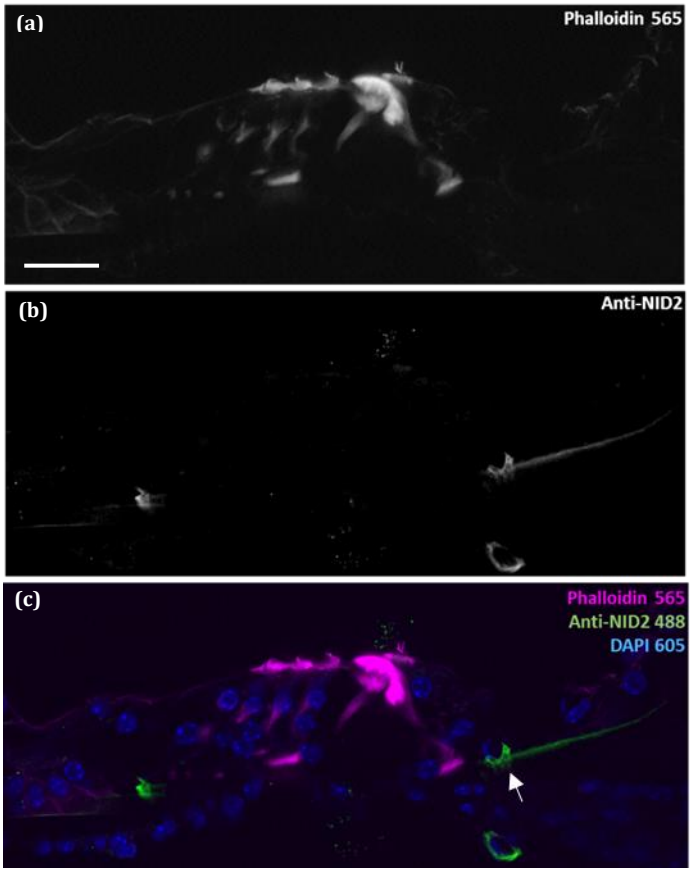

Figure 5.13. Protein localisation of Anti-NID2 at the organ of Corti, 40x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-N**ID2**. The white arrow in (c) points to prominent staining (anti-NID2, green) at a region of the epithelial lining between the inner spiral sulcus and spiral limbus. (c) also displays phalloidin staining in magenta, which highlights the apical ends of sensory hair cells. The scale bar in (a) represents 25µm and is consistent to all three panels in this figure.

5.2.4.2 Rho Guanine Nucleotide Exchange Factor 28 (ARHGEF28)

ARHGEF28 was selected for protein localisation and, similarly to *NID2,* contains lead SNPs that are genome-wide significant in both *HDiff* and *HAid* analysis. In addition, the in *HDiff* analysis, two 'independent' lead SNPs are at the *ARHGEF28* locus and are in different 'LD blocks'. In *HDiff*, rs6890164 and rs6453022 are the lead SNPs that reside in the *ARHGEF28* region, Figure 5.14 and Figure 5.16. rs6890164 lies 6177bp upstream of *ARHGEF28* and rs6453022 resides in exon 7. In the *HAid* analysis the lead SNP rs4597943 lies in intron 7 of *ARHGEF28* Figure 5.15 . The two SNPs rs6453022 and rs6890164 have an LD R^2=0.7573 (p-value <0.0001).

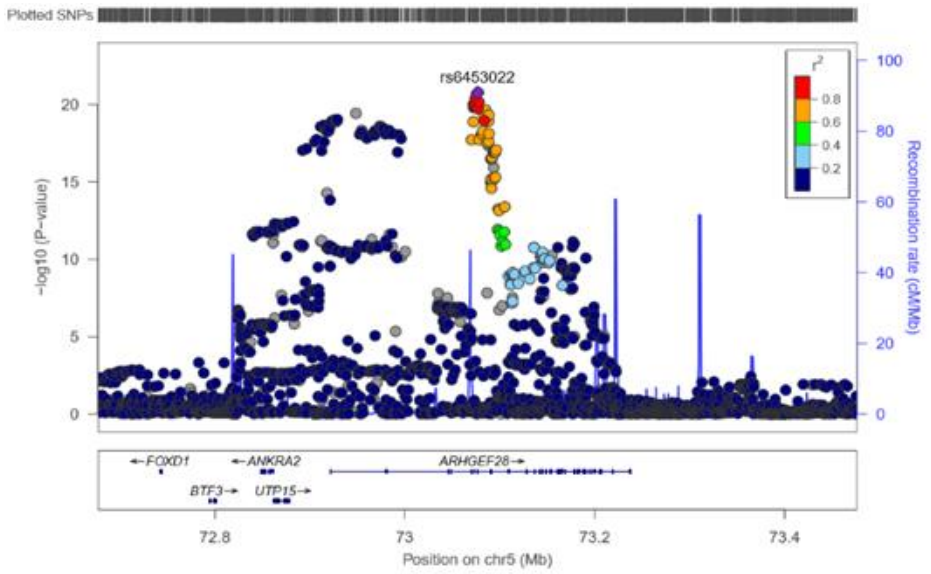

Figure 5.14 Locus plot of *HDiff* lead SNP at the *ARHGEF28* gene transcript

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 50kb +/- regions flanking the position of the lead SNP

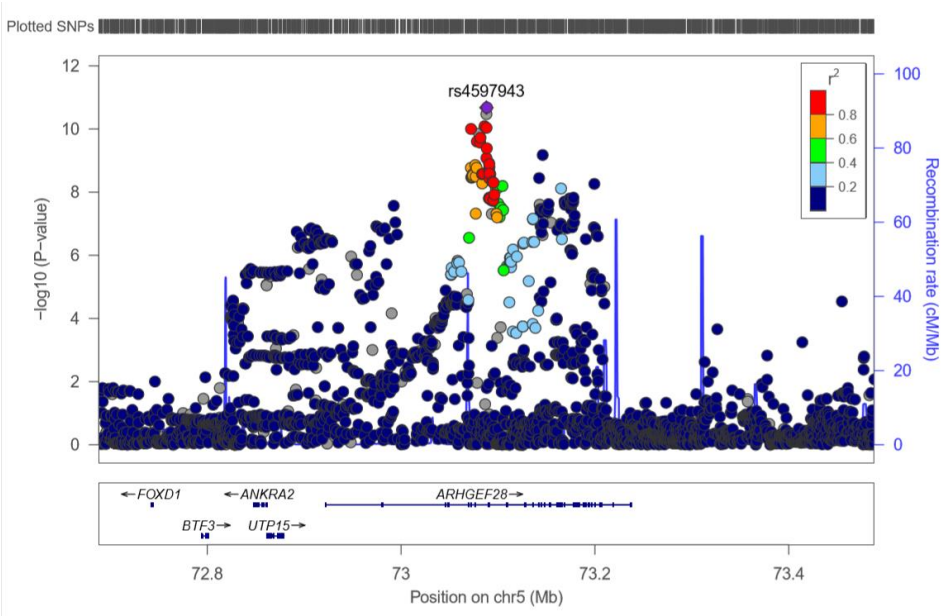

Figure 5.15 Locus plot of *HAid* lead SNP at the *ARHGEF28* gene transcript

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 50kb +/- regions flanking the position of the lead SNP.

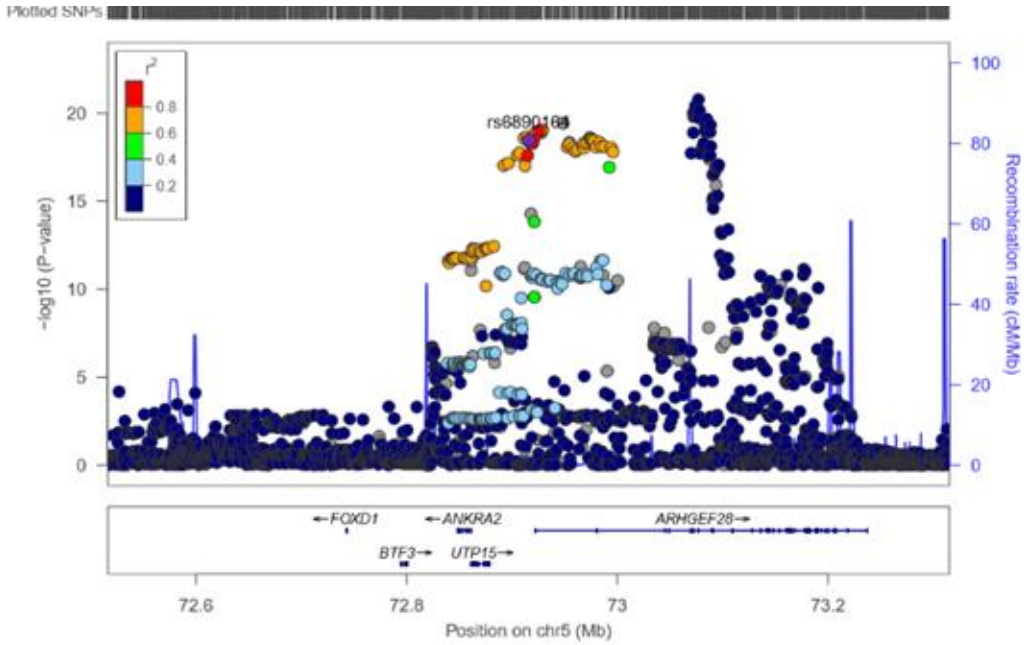

Figure 5.16 Locus plot of *HDiff* SNP in close proximity to the *ARHGEF28* gene transcript

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 50kb +/- regions flanking the position of the lead SNP.

ARHGEF28 codes for Rho Guanine Nucleotide Exchange Factor 28, a member of the Rho guanine nucleotide exchange factor family that activates the Ras-like family of Rho proteins. The exact function has not been established, but when over expressed in neuronal cells, ARHGEF28 has been shown to induce cell rounding and inhibit neurite outgrowth[376]. *ARHGEF28* has not before been associated with auditory function or a pathology, but a number of studies have linked *ARHGEF28* to amyotrophic lateral sclerosis (ALS) and motor neuron disease. ARHGEF28 is a neurofilament mRNA destabilising factor and prior to finding a link between *ARHGEF28* and ALS patients, it was found to have a have a role in forming neuronal cytoplasmic inclusions in spinal motor neurons in mice (p190RhoGEF).

The first link between ARHGEF28 and ALS was found in sample of 7 familial ALS cases, where a frameshift mutation was identified, that is predicted to result in a truncated gene product[377]. Further work on the risk in this family revealed that the *ARHGEF28* mutation may modify the risk via a C9orf72 expansion[378]. A second study to report a link with ALS identified a mutation in ARHGEF28 present in 0.52% of the 380 ALS cases in the Chinese cohort studied[379]. A second study in a Chinese cohort has, more recently, identified a number of variants at distinct locations within the *ARHGEF28* gene which are hypothesised to confer or reduce risk of ALS development[380]. The latest link between *ARHGEF28* and risk of neurodegenerative conditions was in two unrelated Charcot-Marie-Tooth/hereditary motor neuron disease cases[381]. Further to these links with neurodegenerative conditions,

ARHGEF28 has been linked with processes such as cell reorientation[382], cell migration[383] and cell motility[384] in addition to tumour morphology and behaviour[385–387].

Protein localisation of ARHGEF28

At the organ of Corti (Figure 5.17), anti-ARHGEF28 localises at both inner and outer hair cells. Staining is also observed (less prominently) in the stria vascularis. This could be suggestive of a role in the sensory cells; both the inner and outer hair cells. In this section, the rows of outer hair cells have become displaced and so multiple rows are visible (cluster of inner hair cells rather than three, one from each row as seen previously). The 40x image of the organ of Corti(Figure 5.19) also displays this anti-ARHGEF28 staining in the IHCs and OHCs, and staining does not appear to be present in the surrounding supporting cells. In addition to localising at the sensory cells, Anti-ARHGEF28 staining is localised at the spiral ganglion neurons, at the modiolus (Figure 5.18). Prominent staining of Anti-ARHGEF28 is not present however, in the unmyelinated nerve cells that originate from the hair cell synapses (Figure 5.17, Figure 5.19).

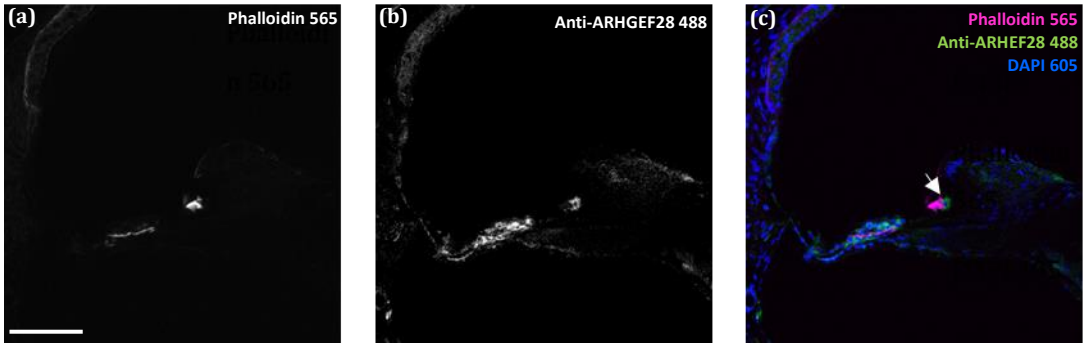

Figure 5.17. Protein localisation of Anti-ARHGEF28 at the organ of Corti, 20x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-ARHGEF28. White arrow in (c) points out anti-ARHGEF28 staining (green) of the inner hair cell. The scale bar in (a) represents 100µm and is consistent for all three images in this figure.

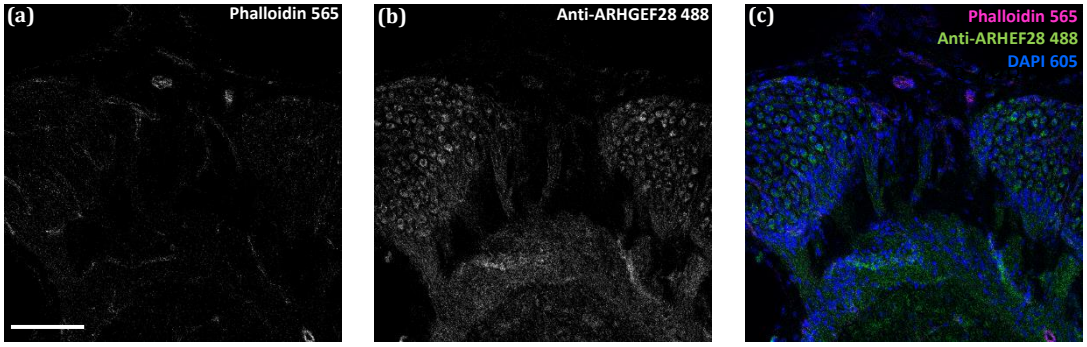

Figure 5.18. Protein localisation of Anti-ARHGEF28 at Spiral ganglion., 20x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-ARHGEF28. Anti-ARHGEF28 staining is seen in white (b) and green (c) at the spiral ganglion neurons at the modiolus. The scale bar in (a) represents 100µm and is consistent for all three images in this figure.

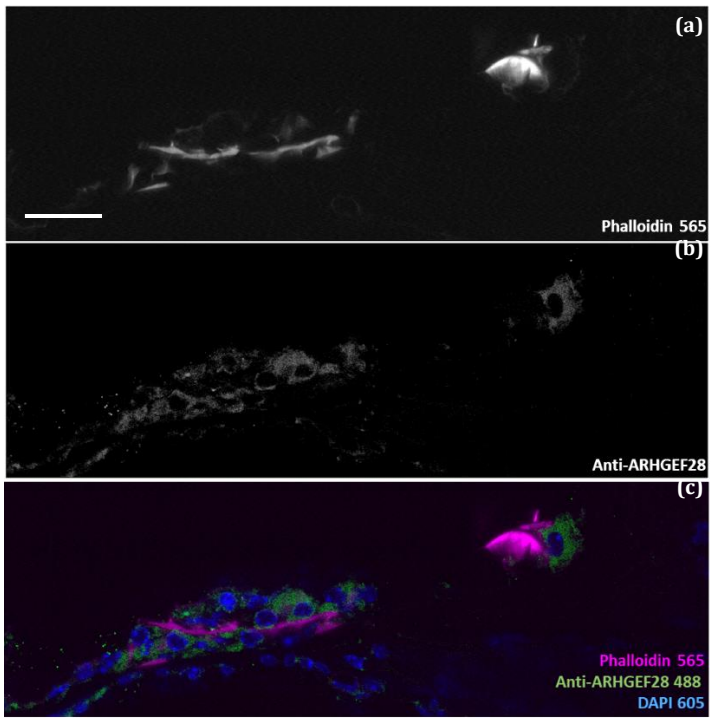

Figure 5.19. Protein localisation of Anti-ARHGEF28 at the organ of Corti, 40x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-ARHGEF28. (c), the composite image, displays phalloidin staining in magenta, which highlights the apical ends of sensory hair cells. Anti-ARHGEF28 staining (green) is present in the cell bodies of the inner and outer hair cells. The scale bar in (a) represents 25µm and is consistent to all three panels in this figure.

5.2.4.3 Clarin-2 (CLRN2)

The lead SNP in the locus containing the *CLRN2* transcript is located 1965bp downstream of the *CLRN2* gene. Within the same LD block as the lead SNP, is the *QDPR* gene transcript (Figure 5.20) *CLRN2* codes for clarin-2, a membrane glycoprotein that is a paralog of *CLRN1*. Mutations in clarin-1 have been shown to cause autosomal recessive Usher syndrome Type-3, a disorder where patients display a sensorineural hearing loss that is present from birth[367,368]. *Ush3a (CLRN1)* expression has been observed in cochlear hair cells and spiral ganglion cells of mice at ages E16, P0, P5 and P10 and so it could be hypothesised that *CLRN2* is also expressed in these cell types.

Additional work has established that the clarin-1 protein localises with microtubules and post-transgolgi vesicles of hair cells, suggestive of a role in neurosensoriepithelia[388]. Lastly, Clarin-$1^{-/-}$ mice exhibit disrupted stereocilia bundles and present an early onset, profound hearing loss[389,390]. Furthermore, as described in the previous section, recent work with *Clrn2*$^{clarinet/clarinet}$ mice has demonstrated an essential role of a functioning *CLRN2* gene for hair bundle integrity and function[349]. As *CLRN1* is expressed in the cochlea and, when mutated, can cause USH3A, *CLRN2* was selected for protein localisation.

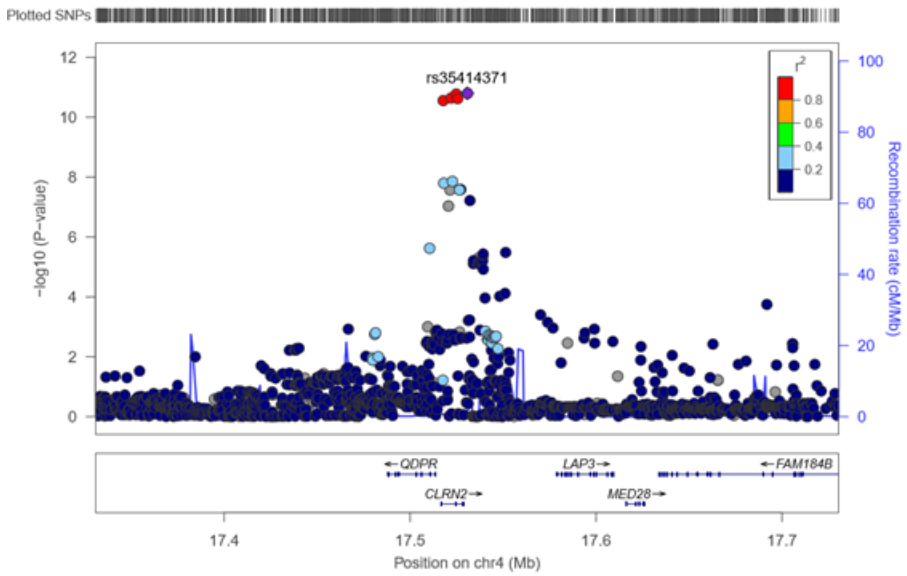

Figure 5.20 Locus plot of the lead SNP in *HDiff* analysis in close proximity to the *CLRN2* gene transcript

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plot displays 50kb +/- regions flanking the position of the lead SNP.

Protein localisation of CLRN2

Under the 20x objective strong anti-CLRN2 staining is observed in outer and inner sensory hair cells and in the stria vascularis, Figure 5.21. Less prominent staining is observed in cells in the basilar membrane. In the 40x image (Figure 5.22), anti-CLRN2 staining at the outer hair cells appears more concentrated at the apical end of the cell.

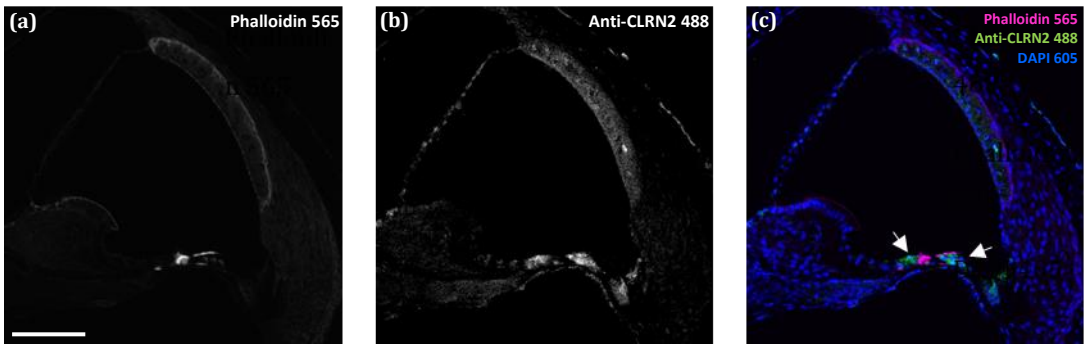

Figure 5.21. Protein localisation of Anti-CLRN2 at the organ of Corti, 20x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-CLRN2. The two arrows in (c) highlight anti-CLRN2 staining of the inner hair cell (arrow directed downward) and the outer hair cells (arrow directed towards the left of the image). The scale bar in (a) represents 100μm and is consistent for all three images in this figure.

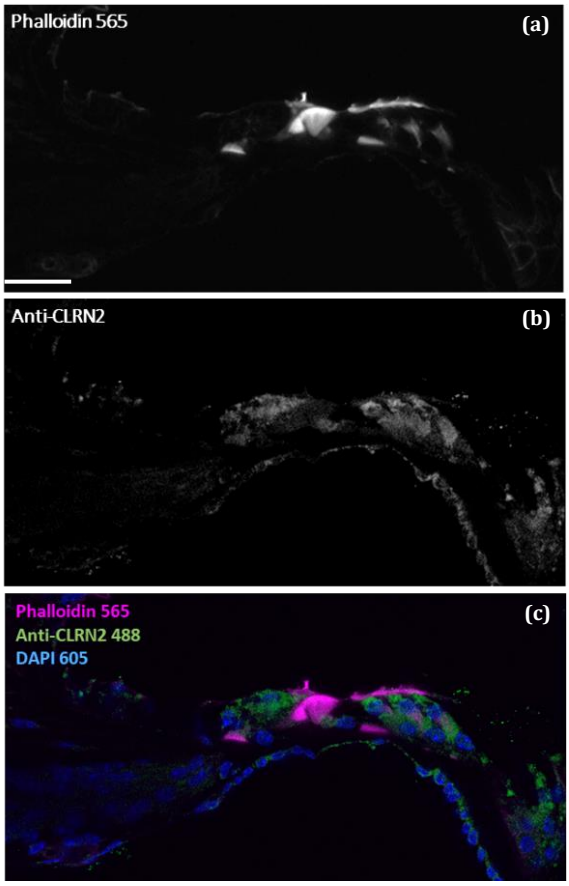

Figure 5.22. Protein localisation of Anti-CLRN2 at the organ of Corti, 40x objective.

Colour labelling in the composite image: Blue, DAPI; Magenta, phalloidin; Green, Anti-CLRN2. (c), the composite image, displays phalloidin staining in magenta, which highlights the apical ends of sensory hair cells. Anti-CLRN2 staining in green is present in the cell bodies of the inner and outer hair cells. The scale bar in (a) represents 25µm and is consistent for all three panels in this figure.

5.2.4.4 Immunohistochemistry discussion

The aim of this section (5.2.4) is to determine the location of protein expression within the cochlea for 3 of the primary gene candidates *NID2*, *ARHGEF28* and *CLRN2.* By determining the location of the protein, a putative pathological site and mechanism can be hypothesised.

Nidogen-2

Nidogen-2 is an established component of the basement membrane complex. Therefore, in the cochlea, it could be hypothesised that nidogen-2 would localise to, and have a structural role, in the basilar membrane. Anti-NID2 staining is not visible at the basilar membrane but is observed at the epithelial lining between the inner spiral sulcus and spiral limbus. The role of this structure is not well established but based on its situation it may function to

support the position and movement of the tectorial membrane. The positioning, support and movement of the tectorial membrane is integral for HCs function; the tectorial membrane has a role in the synchronisation and regulation of feedback from the OHCs and exhibits specialised dynamic mechanics, which create the sub-tectorial space where IHC excitation occurs[391]. Therefore, dysregulation of this structure may lead to a hearing phenotype or damage of the sensory mechanisms. To further explore this hypothesis, tectorial membrane function could be assessed in a nidogen-2 mutant.

At the end of this structure, the strong staining continues into the spike generator region of the cochlea, between the sensory cells and the modiolus of the inner ear. The pattern of this staining is clear in the 40x images of samples stained with anti-NID2 in Figure 5.13. nidogen-1 staining has also been observed in the Reissner's membrane[366] and the osseous spiral lamina yet staining of nidogen-2 was not observed in the samples used in this work. The staining of nidogen-2 is prominent in the lining of blood vessels (Figure 5.12) as hypothesised; previous immunohistochemistry has identified nidogen-2 staining around blood vessels in cardiac muscle and previous staining of nidogen-1 in the cochlea[366]

Rho guanine nucleotide exchange factor 28

No previous studies have tested for ARHGEF28 expression in the cochlea. ARHGEF28 is known to interact with the Ras-like family of Rho proteins and is a neurofilament mRNA destabilising factor. Here anti-ARHGEF28 staining localises to the sensory HCs (inner and outer) and seemingly lines the spiral ganglion nerve fibres. As a neurofilament mRNA destabilising factor, it could be hypothesised that the staining is present in the cytoplasm of the neurons and that the role of the protein may be in the formation or maintenance of the neuronal cytoskeleton.

Based on the location of staining in these samples, it is possible that the location and thus function of ARHGEF28 may not be restricted to the cochlea. Furthermore, in these samples it is not possible to determine which of the anti-ARHGEF28-stained nerve fibres are type I or type II. The function and ARHGEF28-based pathology within this structure could be revealed further by confirming whether *ARHGEF28* is expressed in both or in either of the nerve fibre types.

Staining in the sensory cells is also present, yet it is not known whether the role of ARHGEF28 in these cells is distinct from its role in the spiral ganglion structure. There are multiple splice forms of the transcript and so it is even possible that the resulting protein isoforms could have distinct roles between auditory structures. There are two 'independent' associations within the *ARHGEF28* region in the *HDiff* analysis, represented

by two lead SNPs in distinct blocks of LD; this may even signify multiple functional regions within the transcript that contain variants with distinct roles.

Clarin-2

Clarin-2 (coded for by *CLRN2*) was selected as a candidate for protein localisation as mutations in its paralogue, *CLRN1* are known to cause Ush3a syndrome[368]. In addition, more recently, the *Clrn2*$^{clarinet/clarinet}$ mouse was shown to display a progressive sensorineural hearing loss caused by the failure to maintain HC stereocilia bundle integrity and function[349]. Based on this mechanism, it is hypothesised that clarin-2 is expressed in the HCs and HC stereocilia.

Staining of anti-CLRN2 is observed in the cochlear sensory HCs of p30 mice, as hypothesised. In these samples there is also staining in the stria vascularis, yet there is currently no hypothesised role for clarin-2 in this structure. The function of the stria vascularis is to generate the potassium-rich endolymph of the scala media[392]. The Clarin gene family are integral membrane glycoproteins[68] and so could have a role in this epithelial function. To confirm whether Clarin-2 functions in the stria vascularis, expression could be tested in a *Clrn2*$^{clarinet/clarinet}$ using a consistent protocol and set of antibodies. If expression was not present, this would confirm that the clarin-2 protein is expressed in the stria vascularis structure. To understand the functional interaction between clarin-1 and clarin-2, expression of clarin-2 could be repeated at the same developmental stage as has been tested with clarin-1. Based on the knowledge of Clarin-1 being critical for development of the sensory mechanisms and of clarin-2 maintaining this function, the timepoint at which clarin-1 function becomes redundant could be determined.

Based on the location of the three proteins tested in the adult mouse cochlear tissue, the role of each protein in auditory function can be hypothesised. There are however some limitations to the methods that were used. Firstly, protein localisation is only assessed in the cochlea, one structure within the auditory system. Although staining is observed for all three proteins in this structure, the presence of the protein does not confirm the site of a pathology. Similarly, for all three proteins, the staining is not localised to just one cell type or one individual mechanism. Secondly, the location of the proteins within the structure is estimated based on the structure of the tissue as visualised with DAPI. Other than phalloidin, no other markers are used. To more accurately determine protein location within the structure, further cell-specific markers could be used such as Myosin 7a to mark the outline of the sensory hair cells or GFAP as a marker of glial cells.

Moreover, assays could be performed to confirm antibody specificity for anti-*CLRN2* and anti-*ARHGEF28* that were used here as these have not previously been used in published work. Alternatively, knock-out mouse models for these two genes could be used as a control model to test for non-specific binding. Lastly, the samples are from p28-p30 mice. At this stage adult mice have a fully developed auditory system and thus staining reveals an adult gene-expression profile (rather than a developmental or ageing profile). It would therefore be of value to evaluate gene-expression profiles in aged tissue as the phenotype of interest is of a late-onset, progressive nature.

Chapter 6 - Discussion

Prior to this study, little was known about the genetic component of ARHI, or whether genetic risk factors contribute to ARHI in the general population. The aim of this thesis was to contribute to the knowledge of ARHI genetic risk factors, and subsequently explore what these risk factors can tell us about the underlying biological factors that cause auditory dysfunction in the general population. This final chapter is a discussion of how the work presented in the results chapters has contributed towards answering some of the fundamental questions in this field, followed by a summary of three broad limitations of the study, and lastly some examples of how the findings could be used as a platform for future work.

Summary of main findings

Firstly, for many complex traits, GWAS has proven to be a valuable tool for successfully detecting associated genomic loci[219]. However, in the case of ARHI, previous GWAS had collectively resulted in a limited number of significant associations, and so it was undetermined as to whether GWAS is a suitable method for identifying ARHI risk variants. As a possible reason for these limited findings, Fransen, 2015 speculated that the ARHI 'phenotype depends on the aggregated effect of a large number of SNPs, of which the individual effects are undetectable in a modestly powered GWAS'[38]. The work in this thesis is the first comprehensive study with the means to test this hypothesis, as this can only be tested using a sample with an adequate sample size (and thus statistical power), such as the UKBB.

The principle finding in this thesis is that over 2,000 SNPs are significantly associated with the ARHI-related traits *HDiff* and *HAid*. This amounted to 44 independent genomic loci, increasing the number of known ARHI-associated loci 9-fold. Multiple common genetic variants (MAF>0.01) are associated with the two traits, resembling polygenic inheritance and a strong indication that ARHI has a heritable component. These findings establish the first evidence for the polygenic nature of common hearing impairment for older individuals in the British population.

Secondly, the true heritable component of ARHI has been contested, as previous ARHI heritability estimates were performed using a variety of methods, cohorts and ARHI-surrogate phenotypes, which has resulted in a range of estimates between 0-70%[38,92,106–114]. The range in these estimates, along with the limited number of significant associations identified in previous ARHI GWAS, lead to speculation as to whether ARHI heritability has largely been overestimated. The datasets used in Chapter 4 are the largest samples used to

calculate an ARHI heritability estimate, by an order of magnitude, and therefore has substantially greater statistical power and includes a greater proportion of the population than previous estimates. It therefore has the potential to provide a more comprehensive estimate for the proportion of variance in ARHI based on genetic risk.

In this study SNP heritability (h^2_g) is estimated as 0.19 for *HDiff* and 0.13 for *HAid*. These estimates are lower than some previous estimates, such as those calculated based on PCs derived from PTA thresholds (70%[102]), but are in line with the TwinsUK estimate when SNR score-derived phenotype was used; 25% (adjusted for age)[101]. It could be argued that this SNR-derived phenotype is more consistent with the phenotype used in this study, as both measure hearing ability in background noise and both analyses were adjusted for age. These estimates are of great value to the field as (i) there are few large-scale estimates on the heritability of self-reported hearing ability and (ii) this sample was not recruited based on participant hearing ability.

These *HDiff* and *HAid* heritability estimates and the genetic association analyses reveal that a sizeable proportion of the heritability, and thus the genetic risk of these two traits, is due to common variants. This is in contrast to single gene variants or alterations with large effect sizes, such as those seen in many congenital forms of HL as introduced in Chapter 1. As the estimates calculated in this work are 'SNP heritability' estimates, they reflect the proportion of variance based on the genotyped SNPs included in the association analysis model only. A minor allele frequency of >0.01 was implemented prior to the *HDiff* and *HAid* association analyses, and so this estimate is based on common variants rather than multiple higher effect rare alleles. SNP heritability only considers additive effects of the SNPs selected for input to the calculation and so SNP heritability estimates are generally lower by comparison, as less types of effect (for example, dominant or recessive effects) and the effects of less variants (only those specified in the model), are considered in the calculation. Furthermore, recent work has demonstrated that effect of rare variants, only identified by genome-sequencing, may actually account for much of this 'missing heritability'[393].

A significant factor for ARHI GWAS being 'modestly powered' in work published prior to this study, is that seven of the eight ARHI-GWAS used audiometric data to assess hearing ability[38,92,107,109,111,112,114]. While PTA is the gold-standard method to assess hearing ability in clinic, it requires trained staff, expensive specialist equipment and is relatively time consuming. Collectively, these factors render it unsuitable for large-scale data collection that is required for high powered population studies. One ARHI GWAS had made use of alternative phenotypes, (Hoffman, 2016), where two genetic loci were significantly

associated with phenotypes derived from ICD codes from electronic health records collected in the USA. The findings were replicated using a preliminary release of UKBB genetic data[113].

The work in this thesis is highly distinctive in that a key aim was to search for an alternative way to assess the ARHI status of participants on a large scale, focusing on the full release of the UKBB data. The aim was to determine whether the data could be used to derive a surrogate phenotype that both accurately represents ARHI in the sample and that was applicable for use as a phenotype for genetic association analysis. Work focused on the UKBB SIN test and the hearing-related questions included in the UKBB health and lifestyle questionnaire. As the largest collection of SIN data on one population cohort, the UKBB data also has a potential to set standards for future SIN testing. The data challenge previous standards for SIN testing and accepted characteristics of ARHI, which is the trait that most SIN tests are devised to assess.

Firstly, the prevalence of hearing difficulty in the UKBB as assessed by SRT score categories defined by Smits et al[88], is lower than expected for a sample of this age range. These findings contest the validity of either the previous population prevalence estimates, or the method of classification used to determine hearing ability based on UKBB SRT scores. Secondly, the results contest pre-established risk factors; no significant difference was observed between male and female SRT score distributions, contrary to previous data[272,273] and the relationship between age and SRT score was modest (r^2_s = 0.238, p<2.2e-16) compared to previous reports[96]. In addition, the data suggest that bilateral hearing impairment is not as prevalent as previous estimates with PTA data have suggested; the correlation coefficient of SRT_L and SRT_R within the sample is r^2s=0.358, p-value< 2.2e-16. Alternatively, this finding may reflect a low test-retest reliability within participants, as supported by the longitudinal data analysis.

These results are especially valuable as they challenge current understanding of ARHI epidemiology and risk factors and bring into question whether previous measures of common adult hearing impairment, such as PTA scores or previous SIN tests, have accurately captured ARHI in the population. The results suggest that the correlation with age is not as high as previously estimated, that male sex may not be a significant risk factor and the prevalence of a bilateral deterioration is not as prevalent as previously thought. The work is also of importance to the field as it reveals a number of limitations of the UKBB SIN test; a low test-retest performance, significant differences in SRT score distributions between assessment centres and inconsistent protocol adaptations over time. Therefore,

while the data may challenge current understanding and progress knowledge of ARHI epidemiology, the limitations uncovered and supported by this work, regarding the reliability and sensitivity of the SIN test, ought to be recognised when establishing new standards for ARHI assessment.

This is the first study to evaluate whether questionnaire data can be used to create a surrogate measure of ARHI for use in GWAS. No previous ARHI GWAS has used phenotypes derived from questionnaire responses, but in recent years this approach has been successfully used in genetic association studies for numerous complex traits[225,394–396]. This study has shown for the first time that by using a combination of responses from two questions regarding two key symptoms of ARHI; (i) hearing difficulty and (ii) difficulty in the presence of background noise, a qualitative phenotype measure can be derived that meets the expected population prevalence of ARHI and which displays expected trends such as progression with age and prevalence differences in male and female subjects[85–88].

These findings are an important contribution to the field of ARHI epidemiology as questionnaires are a quick and accessible way of collecting data. There is no requirement for expensive equipment, trained staff or for the individual to attend a clinic visit (a challenging task for some ageing patients). Therefore, a larger sample can be collected, and greater statistical power achieved, at a fraction of the cost of audiometric data collection. Similarly, collecting data on treatments such as hearing aids, negates the need to align clinical databases, a process which can be lengthy and complex due to ethical considerations and computational demands. In addition, longitudinal data can be collected remotely either online or by postal questionnaires.

The two most recent and most successful (when defined by quantity of significant associations) ARHI GWAS attempts (Hoffman 2016[113] and work presented in this thesis) use a combination of phenotypes that move away from the traditional approach of using phenotypes derived from audiometric data[38,92,107,109,111,112,114] for GWAS. The results of both these studies support the use of large data samples that result in increased statistical power compared to smaller samples with more detailed phenotyping[38,92,107,109,111,112,114]. These studies therefore pave the way for future work to use similar techniques, enabling progress which had not been feasible with the use of previous methods.

Previous to this work, a number of traits had been associated with ARHI, mainly based on findings from epidemiological studies. Whether these associations are due to common risk factors, common biological pathologies and, or cause and effect relationships between traits, was largely undetermined[123,140,165,179–181], as described in detail in section 1.3.3. The

genetic correlation analysis presented in Chapter 4 demonstrates for the first time that a number of these traits (such as tinnitus, depression and pain symptoms) have significant genetic correlations with the *HDiff* phenotype when using the LDSC[251] method. Prior to this work, there was no evidence of a genetic correlation between these traits, or that there were any common biological underpinnings between the traits. By uncovering such correlations, we can better understand the biology of ARHI, the correlated traits, and the relationship between them. This work could be built upon by identifying common genetic risk factors and thus common pathological pathways. Ultimately, cause and effect relationships between traits can be investigated using techniques such as Mendelian randomisation[397] or a recently developed latent causal variable (LCV) model[398] that can infer causality.

Perhaps most importantly, three quarters of the associated loci are novel associations with HL or hearing function. This represents a substantial increase in the number of ARHI risk gene candidates. They also greatly progress the knowledge of the type of mechanisms and structures that are affected in common hearing impairment. Prior to this work, little was known of these mechanisms, yet together, these candidate genes have roles in all four structures implicated in Schuknecht's subtypes, meaning that the pathology of common hearing impairment spans multiple cell types in the auditory system. This substantiates the hypothesis that ARHI likely has a mixed pathology within the population and possibly within individuals (due to the relatively small effects of individual variants). Numerous genes at associated loci are known to be expressed in sensory cells such as *CLRN2*, while others have predicted neuronal functions such as *ARHGEF28*, metabolic functions such as *TYR*, and possibly mechanical functions such as *NID2*.

Three of the associations with novel hearing genes have been validated by expression analysis of *ARHGEF28*, *NID2* and *CLRN2* (Chapter 5). Further to this, work regarding *CLRN2* is a key example of how an association can be validated and further investigated by functional analysis. Alongside this study revealing a significant association with *HDiff* at the *CLRN2* locus and subsequent protein localisation in sensory cells in adult mice, work from a forward genetic screen in mice uncovered a role of *Clrn2* in hair cell stereocilia maintenance and function[369]. While a plausible ARHI gene candidate due to *CLRN1* mutations causing USH3[68,368,389,390], no previous links had been made between *CLRN2* and function or pathologies. Collectively, the recent findings regarding CLRN2 are evidence that it likely has a role in maintenance rather than development, and so may well encode risk variants that contribute to 'case' phenotypes in *HDiff*.

The remaining quarter of candidate genes identified have been linked genes to human hearing pathologies that previously had known functions in the auditory system; *CTBP2, TUB, SYNJ2* and *SPTBN1*. The link with *CTBP2* is strong evidence of the reliability of this work; a splice variant of *CTBP2* codes for ribeye, a main component of the HC synaptic ribbon. CTBP2 is a well-established component of this structure and antibodies against the protein are often used as a marker for HCs in localisation analysis. The association here between *HDiff*, *HAid* and *CTBP2* variants also supports the theory of synaptopathy being a crucial step in the degradation of hearing in ARHI and noise-induced hearing loss something that was previously overlooked and, or unidentified[36,399].

At a number of these loci there is strong evidence to support the reliability of the association findings; six of the lead SNPs reside in, or in close proximity to, genes with links to different forms of hearing loss; *EYA4, BAIAP2L2, TRIOBP, ILDR1, LMX1A* and *CDH23*. Furthermore, two of these genes; *TRIOBP* and *ILDR1,* have previous links to ARHI-related phenotypes. This supports the authenticity of the findings, and even challenges current understanding of the specific function of certain genes within the auditory system. For example, *LMX1A* is a LIM homeodomain-containing transcription factor that has multiple roles in development and, specifically in the inner ear, has long been known to have a vital role in early development and structure formation[400]. The finding of an association between *LMX1A* and common, progressive HL traits, is suggestive of a maintenance role, beyond early development. This supports a recent study in mice where heterozygote mutations in the homeodomain of *LMX1A* do not result in severe developmental defects, but result in insufficient cochlear (and sometimes vestibular) function and maintenance[401].

Each candidate gene is of importance with respect to understanding the development of ARHI. For genes where the function is known, and even where the function in the auditory system is known, hypotheses of the meaning of these new associations can be readily proposed. However, for genes which have no previous link to hearing and where knowledge of gene function is limited, it is more difficult to derive such hypotheses. These associations may however be the most important as they could potentially provide the greatest advancements in our knowledge. *ARHGEF28* for example, is highly associated with both *HDiff* and *HAid* yet there is no previous link to hearing function or a comprehensive understanding of how the gene functions in other tissues. By further exploring the function of such gene candidates, it may be possible to describe previously unknown biological mechanisms that are involved in auditory function or, more broadly, in multiple tissues affecting multiple conditions. Furthermore, the list of candidates that are novel to hearing function may actually be important findings in terms of other hearing conditions. This work

provides a list of possible candidates that could be used for screening to identify causes of other forms of hearing loss, where the pathology is currently unknown.

Limitations and Future Work

The work presented in this thesis has greatly improved our knowledge of the genetic risk of ARHI, and the suitability of multiple methods for studying the genetic risk of the condition. There are, however, a number of limitations that ought to be considered when interpreting the findings and also when performing work that builds on the results of this study. As the analysis-specific limitations have been addressed in individual chapter discussions, three broad three limitations that relate to the findings collectively, are summarised below.

Firstly, in GWAS studies, it is an established practice to replicate the association analysis in an independent population sample to eliminate the possibility of the findings being due to cohort or study-specific effects, as described in Chapter 1. A successful replication is a form of validation of the findings, and signifies that the results may be applicable to the wider population[231]. As discussed in Chapter 4, the replication meta-analysis resulted in three Bonferroni-corrected significant lead SNP replications, and just two of the leading SNPs highlighted in previous ARHI GWAS were replicated in the *HDiff* and *HAid* discovery analysis.

The reason for this relatively low replication rate could be due to the heterogeneity between this study and those previous to it, and the relatively limited size of the replication sample. The replication meta-analysis and the Hoffman 2016[113] GWAS sample (the only samples with significant replication statistics) were a magnitude smaller than the UKBB white British sample, and all other previous samples were two magnitudes smaller. As more large-scale datasets become available, replication cohorts are more likely to have sufficient statistical power to replicate findings. For example, 23andMe (n>500,000 European descent[402]) includes questionnaire-derived hearing data.

Until such data are analysed and, or become publicly available, the *HDiff* and *HAid* association results can be interpreted and assessed for reliability in other ways. Candidate genes from this work can be assessed for reliability by determining whether they have previously been implicated in other forms of human hearing loss, been identified as pathogenic markers in mouse screens, have a plausible biological mechanism such as confirmed expression in cochlear sensory structures, or have a role in neuronal maintenance and function. All of these situations would provide confidence for the relevant associations found in this work, as it suggests a role for the gene candidate in auditory function.

A second but related limitation of this work, is that by using the 'white British' and 'white non-British' populations, the findings are not necessarily applicable to the wider population. Population groups that were not within the PC components used to derive these population classifications may harbour distinct ARHI genetic risk loci, or distinct functional variants at the same loci as identified here. This limitation is common to the majority of published GWAS studies. There is a bias towards data collection on European samples and thus a bias as to whether results are applicable to diverse populations; in 2017, 88% of the published GWAS used samples of European Ancestry. Almost two thirds of these studies were from US, UK and Icelandic populations[403]. As a consequence of this ethnic bias, imputation reference maps and genotyping platforms that are accurate for samples of non-European descent are far smaller in size and number than those for populations of European descent[404], further limiting the diversity of studies and wider applicability of findings.

As discussed in Chapter 1, differences in ARHI prevalence have been observed between different populations[95–98], meaning that the limitation outlined above is likely to affect the applicability of these findings to the wider population. The imputed UKBB data used in this thesis, was imputed centrally by UKBB, using the HRC reference panel and the UK 10K and 1000G reference panels. These were derived predominantly using European data and therefore are not suitable for multi-ethnic analysis, even though large subsets of other populations are present in the UKBB dataset. Imputation using appropriate reference maps and, or sequencing would be valuable to counteract this limitation and to explore whether differences in ARHI prevalence and progression between populations are based on genetic factors.

A third limitation of the work presented in this thesis is the interaction of risk factors such as sex-specific effects and environmental exposures were not comprehensively explored due to time restrictions. As discussed in Chapter 1, multiple environmental risk factors have been associated with ARHI such as chemical exposures, alcohol intake, smoking and noise exposure[120–124]. The UKBB data does include a subset of data related to these exposures, but this was not explored here due to time limitations. The influence of these factors on ARHI development are important to understand, due to gene environment interactions in the context of an individual's risk. To adjust for this, individuals' genetic sex was included as a covariate in the genetic association analyses, but analysis was not stratified by sex. Stratifying the analysis by sex may have revealed sex-specific genetic risk factors, and thus progressed the knowledge of sex-specific risk factors, but it would have required halving the sample size of each analyses and thus vastly reduced the statistical power to identify significant associations.

The work described in this thesis provides a platform for future work to be carried out to furtherthe findings. Future work will also provide an opportunity to address the limitations that have been described above if certain data and, or analysis tools are accessible. The UKBB data resource offered a unique opportunity to achieve the aims set out in this thesis. The dataset is the largest, most comprehensive dataset to include hearing-related and genetic data fields and that is publicly available for research. This study has demonstrated the value of the UKBB dataset as it has substantially progressed the field of ARHI genetics, and it is likely that the potential of the resource to progress the field has not been fully realised, as discussed below.

The broad and ongoing data collection at UKBB means that its potential for use in auditory research has not been exhausted. ARHI is understood to be a heterogeneous condition, pathological mechanisms and the genetic variants underlying such mechanisms are likely to vary within the population. This can result in a loss of power when using one method to define the condition. By subsetting the sample into groups based on hypothesised pathology, greater power may be gained in order to detect pathology-specific variants. For example, a comparison could be performed between individuals that have hearing difficulty in background noise but report no general difficulty (UKBB Field ID 2247.0.0), with those that have general hearing difficulties but not in the presence of background noise (UKBB Field ID 2257.0.0). This topic could be explored further by using several approaches.

Firstly, by using the cognition data in conjunction with the hearing data, to subset participants by theorised pathologies. The UKBB data collection includes seven cognitive function tests; prospective memory (UKBB Field ID 20018), pairs matching (UKBB Field ID 20197), fluid intelligence (UKBB Field ID 20016), reaction time (various Field IDs), symbol digit substitution (UKBB Field ID 20159), trail making A and B (various Field ID). Studies have already used these to search for links between cognition and hearing in the sample, but none has yet incorporated the effects of genetic risk[122,162]. In addition, the UKBB SIN results have been used to observe a relationship between poor SRT_B score and increased levels of grey matter in whole brain and in predicted functional networks on a subset of the UKBB population (N=8,701, mean SRT_B = -6d_B, 61% showing normal hearing)[238]. This could be investigated further by incorporating genetic risk into the model, in an attempt to quantify the genetic interactions resulting in these changes.

Secondly, non-genetic risk factors for ARHI can be explored with the UKBB data, as previously demonstrated by a number of epidemiological studies[122,141,162]. These factors are also yet to be studied in conjunction with the UKBB genetic data. Data fields such as noise

exposure (Field IDs 4836 and 4825) and antibiotic use (various Field IDs) could be used in conjunction with hearing question response data to identify gene x environment interactions and identify variants that have a protective effect, such as those with roles in damage response mechanisms.

Thirdly, the recruitment of participants was devised such that a recruitment bias and recall bias would be limited. The age of participants at recruitment means that longitudinal data can be collected as data is, and will continue to be, gathered regarding new diagnoses and the progression of numerous conditions. Therefore, the onset of hearing impairment for a number of participants will be recorded, as will the onset of other related conditions such as cognitive decline. This could increase statistical power to detect risk variants as participants can be subset not only based on symptoms, but also of age at symptom onset, permitting a better understanding of the interactions between different conditions. Current plans to link the UKBB resource to public health databases means that prospective analysis will also be permitted.

In addition to phenotype data releases, future UKBB genetic data releases will provide an opportunity to further the work presented in this thesis. Currently, the UKBB is in the process of releasing WES data on a sample of 50k participants. As genotyped and imputed data included in the GWAS encompasses only a fraction of the genome, such data could be used to fine map and identify the true functional variant(s) in the six regions where the lead SNP here resides in a gene exon; rs36062310, *KLHDC7B;* rs6453022, *ARHGEF28;* rs9493627, *EYA4;* rs5756795, *TRIOBP* ; rs143282422, *CDH23;* rs12552, *OLFM4.* Following this, WES and WGS are to be released on samples from the whole UKBB cohort, which could be utilised for both fine mapping all regions of association and for detecting association between rare variants and the two traits studied here. Prior to this data release, WES from samples such as used in Girotto 2018(N=156)[405] or from the TwinsUK sample (N=2000, UK10K; N=2377, HLI next-gen sequenced data)[406] could be used for fine mapping regions of interest. Identification of very rare alleles from WGS will however require samples that are magnitudes larger than these samples.

Further to using the UKBB data to progress the findings in this work, there are many avenues to explore to continue the post-GWAS analysis. In much of the post-GWAS analysis presented here, the emphasis was on protein-coding genes as the aim of Chapter 5 was to select candidates for immunohistochemistry analysis. However, in the context of following up genetic variant associations, non-coding regions must also be included; ENCODE (https://www.encodeproject.org/) predicts that only 1% of the genome comprises of gene

transcripts and it is widely recognised that variants involved with complex traits are enriched in regulatory regions. Numerous tools are being developed and improved to study regulatory elements in this context such as transcription factor binding site databases such as ChIPBase (http://rna.sysu.edu.cn/chipbase/) or eQTL mapping via GTEx (https://www.gtexportal.org/home/). The GTEx v8 will make available expression data in a greater number of tissues including a number of distinct brain regions, albeit not including the peripheral auditory system[364].

When studying regulatory elements, epigenetic features also ought to be assessed; GWAS findings currently only explain a fraction of trait heritability estimates, and epigenetic features are hypothesised to explain a proportion of this missing heritability for many complex traits[293]. In addition, recent work has demonstrated that for traits such as BMI and cigarette consumption, studying epigenetic variation under a Bayesian statistical model proposed by Banos[407], phenotype prediction is improved and novel associations can be identified. An epigenome-wide association scan (EWAS) exploring the association between DNA methylation levels in whole blood samples and ARHI identified and replicated high associations in promoter regions of *TCF25* and *POLE*[222]. Further work could include EWAS on auditory tissue in a larger sample.

A pathogenic variant detected via association analysis may act by influencing the behaviour or function of a downstream target. In this context, it is more likely that the downstream target will have a function in auditory mechanisms and therefore, pathway analysis is an important step in deciphering the biological relevance of associated genetic variants. There is however a relative lack of expression data from auditory tissue samples in publicly available annotation tools; between the two main data repositories for expression data in the inner ear, gEAR (https://umgear.org/, 13 datasets available) and SHIELD (https://shield.hms.harvard.edu/, 5 datasets available), there is only one study that contains expression data on aged samples (mice >P40[222]). The volume of relevant data is increasing however, for example the recent RNA-sequencing screen on adult mouse stria tissue[408] increases the amount of relevant data, as it was performed on adult rather than pup samples and was not restricted to the inner ear sensory cells.

While the main aim of identifying ARHI-risk loci is to gain insights into the pathology, methods are being devised that permit trait prediction based on genetic association results. For example, polygenic risk scores (PRS) can be calculated for individuals in a 'target sample' (such as clinic patients) based on genetic effects observed from a 'training sample' (here being the *HDiff* summary statistics)[409]. Although PRSs are predominantly not yet valid

for clinical use, the current methods are a step towards diagnosis and treatment for complex traits based on individual genetic risk and can be used to compare genetic risk of multiple traits. In the short-term, the understanding of ARHI pathology could benefit greatly from further analysis in this area due to the relatively little amount known about ARHI pathology and the great number of significantly correlated traits. A long-term goal of analysis such as this would be to contribute to multi-trait treatments or prevention strategies.

In vitro and in vivo work will be an essential component of the post-GWAS analysis. The suitability of the mouse as a model for ARHI genetics means that much of this work could be conducted with mice. For example, further expression analysis such as presented in Chapter 5, but for a greater number of candidate genes. This would support findings from the association analysis, and indicate which structures are affected in gene-specific pathologies. Furthermore, as discussed in Chapter 5, expression profiles under different stress conditions and time points could lead to hypotheses of gene function.

Forward genetic screens have already been used in this context, in the case of *Clrn2*, detailed above, and could be conducted with a number of the candidate genes presented in Chapter 5. Where forward genetic screens are not feasible due to facilities or lethal effects of gene disruption or deletion in early development, alternative methods such as gene knock-down could be used to assess gene function in aged, wild-type mice. Using targeted editing techniques such as CRISPR-Cas9, it may be possible to deduce which functional gene regions correspond to varied phenotypes resulting from defects in the same gene. *EYA4*, as discussed in Chapter 5, is an example of this; multiple forms of hearing impairment are observed and are understood to be caused by variation at distinct regions within the gene[204,288,289].

In addition to forward genetic approaches, reverse genetic approaches could also be invaluable in the assessment of ARHI GWAS candidate genes. Reverse genetic approaches involve disrupting a gene, most commonly by knocking out or silencing the gene, and studying the subsequent phenotype. Where a phenotype relates to a trait of interest, hypotheses can be made about the function of the knocked-out or silenced gene. The International Mouse Phenotyping Consortium[410] (IMPC) is a large-scale programme that aims to generate knock-out mouse models for each of the ~20,000 genes in the mouse genome. Aligning IMPC data with ARHI GWAS candidate genes could aid the validation of a genetic association, identification of a candidate gene within an associated region, and the function of a gene in the auditory system.

While several of the candidate genes in this study have been validated, there is a proportion of genes that are relatively under-studied and less well characterised. This leads to a bias towards functional annotation of well characterised genes as there is more, and indeed more relevant, data available. In the study of hearing function, the untargeted approach of the IMPC will provide an opportunity to validate and further explore the roles of genes that are less well characterised. It is also worth noting that several of the associated regions may represent a gain of function variant. In these examples, a knock-out or silenced mouse model generated via the IMPC would not provide a suitable model to study the effects of the SNP on hearing function. However, if the gene is involved in normal hearing function, the mouse model will display a phenotype. The observation of a phenotype could therefore support the finding of an association with the candidate gene, and a validation that the candidate gene is the functional target of the associated SNP in the region.

Concluding statements

To conclude, the work presented in this thesis confirms for the first time the hypothesis that ARHI is a highly polygenic trait. The work demonstrates that the genetic risk of ARHI (in the UKBB population) consists of multiple common polymorphisms with relatively small effect sizes, and that GWAS is a valuable method by which to identify putative ARHI genetic risk loci. The work uncovered the value of self-report measures to classify hearing impairments in large cohort studies, and highlighted limitations of the UKBB SRT data in its current form.

A quarter of the genetic risk loci identified in this study contain genes that are known to cause other forms of HL, and, or that have an established role in hearing function. The remaining risk loci identified resemble novel associations with HL, greatly expanding the knowledge of biological mechanisms involved in common hearing impairment. A better understanding of these mechanisms may lead to more comprehensive methods to diagnose, treat and even prevent forms of HL and associated conditions.

References

1. Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia, A.-S., McNamara, J.O., and Williams, M.S. (2001). Neuroscience, 2nd edition (Sinauer Associates).

2. Anthwal, N., and Thompson, H. (2016). The development of the mammalian outer and middle ear. J. Anat. *228*, 217–232.

3. Bartel-Friedrich, S., and Wulke, C. (2007). Classification and diagnosis of ear malformations. GMS Curr. Top. Otorhinolaryngol. Head Neck Surg. *6*, Doc05.

4. Mansour, S., Magnan, J., Nicolas, K., and Haidar, H. (2018). Otosclerosis. In Middle Ear Diseases, (Cham: Springer International Publishing), pp. 1–83.

5. Pollak, A., Lechowicz, U., Kędra, A., Stawiński, P., Rydzanicz, M., Furmanek, M., Brzozowska, M., Mrówka, M., Skarzyński, H., Skarzyński, P.H., et al. (2016). Novel and De Novo mutations extend association of POU3F4 with distinct clinical and radiological phenotype of hearing loss. PLoS One *11*, e0166618.

6. Khan, S., and Chang, R. (2013). Anatomy of the vestibular system: A review. NeuroRehabilitation *32*, 437–443.

7. Goutman, J.D., Elgoyhen, A.B., and Gómez-Casati, M.E. (2015). Cochlear hair cells: The sound-sensing machines. FEBS Lett. *589*, 3354–3361.

8. Iwasa, K.H., Mizuta, K., Lim, D.J., Benos, D.J., and Tachibana, M. (1994). Amiloride-sensitive channels in marginal cells in the stria vascularis of the guinea pig cochlea. Neurosci. Lett. *172*, 163–166.

9. Chan, D.K., and Chang, K.W. (2014). GJB2-associated hearing loss: Systematic review of worldwide prevalence, genotype, and auditory phenotype. Laryngoscope *124*, E34–E53.

10. Gabriel, H., Kupsch, P., Sudendey, J., Winterhager, E., Jahnke, K., and Lautermann, J. (2001). Mutations in the connexin26/GJB2 gene are the most common event in non-syndromic hearing loss among the German population. Hum. Mutat. *17*, 521–522.

11. Wilcox, E.R., Burton, Q.L., Naz, S., Riazuddin, S., Smith, T.N., Ploplis, B., Belyantseva, I., Ben-Yosef, T., Liburd, N.A., Morell, R.J., et al. (2001). Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. Cell *104*, 165–172.

12. Riazuddin, S., Ahmed, Z.M., Fanning, A.S., Lagziel, A., Kitajiri, S.I., Ramzan, K., Khan, S.N., Chattaraj, P., Friedman, P.L., Anderson, J.M., et al. (2006). Tricellulin is a tight-junction protein necessary for hearing. Am. J. Hum. Genet. *79*, 1040–1051.

13. Goutman, J.D., Elgoyhen, A.B., and Gómez-Casati, M.E. (2015). Cochlear hair cells: The sound-sensing machines. FEBS Lett. *589*, 3354–3361.

14. Musiek, F.E., and Baran, J.A. (2007). The auditory system : anatomy, physiology and clinical correlates. (Boston; Montreal : Pearson), p. 487.

15. Davis, H., and Silverman, S.R. Hearing and deafness. (Rinehart & Winston of Canada Ltd;).

16. Riazuddin, S., Khan, S.N., Ahmed, Z.M., Ghosh, M., Caution, K., Nazli, S., Kabra, M., Zafar, A.U., Chen, K., Naz, S., et al. (2006). Mutations in TRIOBP, which encodes a putative cytoskeletal-organizing protein, are associated with nonsyndromic recessive deafness. Am. J. Hum. Genet. *78*, 137–143.

17. Miyagawa, M., Nishio, S. ya, and Usami, S. ichi (2012). Prevalence and clinical features of hearing loss patients with cdh23 mutations: A large cohort study. PLoS One *7*, e40366.

18. Varga, R., Kelley, P.M., Keats, B.J., Starr, A., Leal, S.M., Cohn, E., and Kimberling, W.J. (2003). Non-syndromic recessive auditory neuropathy is the result of mutations in the otoferlin (OTOF) gene [2]. J. Med. Genet. *40*, 45–50.

19. Iwasa, Y. ichiro, Nishio, S. ya, Sugaya, A., Kataoka, Y., Kanda, Y., Taniguchi, M., Nagai, K., Naito, Y., Ikezono, T., Horie, R., et al. (2019). OTOF mutation analysis with massively parallel DNA sequencing in 2,265 Japanese sensorineural hearing loss patients. PLoS One *14*, e0215932.

20. Nayagam, B.A., Muniak, M.A., and Ryugo, D.K. (2011). The spiral ganglion: Connecting the peripheral and central auditory systems. Hear. Res. *278*, 2–20.

21. Weisz, C., Glowatzki, E., and Fuchs, P. (2009). The postsynaptic function of type II cochlear afferents. Nature *461*, 1126–1129.

22. Brownell, W.E., Bader, C.R., Bertrand, D., and De Ribaupierre, Y. (1985). Evoked mechanical responses of isolated cochlear outer hair cells. Science (80-.). *227*, 194–196.

23. Guinan, J.J. (2006). Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans. Ear Hear. *27*, 589–607.

24. Eybalin, M. (1993). Neurotransmitters and neuromodulators of the mammalian cochlea. Physiol. Rev. *73*, 309–374.

25. Webster, Douglas B; Fay, R.R. (2005). The Mammalian Auditory Pathway: Neurophysiology (Springer-Verlag).

26. Schuknecht, H.F. (1964). Further Observations on the Pathology of Presbycusis. Arch. Otolaryngol. *80*, 369–382.

27. Landegger, L.D., Psaltis, D., and Stankovic, K.M. (2016). Human audiometric thresholds do not predict specific cellular damage in the inner ear. Hear. Res. *335*, 83–93.

28. Wu, P.Z., Liberman, L.D., Bennett, K., de Gruttola, V., O'Malley, J.T., and Liberman, M.C. (2019). Primary Neural Degeneration in the Human Cochlea: Evidence for Hidden Hearing Loss in the Aging Ear. Neuroscience *407*, 8–20.

29. Liberman, M.C., and Kiang, N.Y.S. (1978). Acoustic trauma in cats. Cochlear pathology and auditory-nerve activity. Acta Otolaryngol. *86*, 1–63.

30. McGill, T.J.I., and Schuknecht, H.F. (1976). Human cochlear changes in noise induced hearing loss. Laryngoscope *86*, 1293–1302.

31. Schuknecht, H.F., and Gacek, M.R. (1993). Cochlear pathology in presbycusis. Ann. Otol. Rhinol. Laryngol. *102*, 1–16.

32. Fujioka, M., Okano, H., and Edge, A.S.B. (2015). Manipulating cell fate in the cochlea: A feasible therapy for hearing loss. Trends Neurosci. *38*, 139–144.

33. Johnsson, L.G. (1974). Sequence of degeneration of corti's organ and its first-order neurons. Ann. Otol. Rhinol. Laryngol. *83*, 294–303.

34. Spoendlin, H. (1984). Factors inducing retrograde degeneration of the cochlear nerve. Ann. Otol. Rhinol. Laryngol. Suppl. *112*, 76–82.

35. Liberman, M.C., and Kujawa, S.G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. Hear. Res. *349*, 138–147.

36. Liberman, M.C. (2017). Noise-induced and age-related hearing loss: New perspectives and potential therapies. F1000Research *6*, 927.

37. Suzuki, J., Corfas, G., and Liberman, M.C. (2016). Round-window delivery of neurotrophin 3 regenerates cochlear synapses after acoustic overexposure. Sci. Rep. *6*, srep24907.

38. Fransen, E., Bonneux, S., Corneveaux, J.J., Schrauwen, I., Di Berardino, F., White, C.H., Ohmen, J.D., Van De Heyning, P., Ambrosetti, U., Huentelman, M.J., et al. (2015). Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. Eur. J. Hum. Genet. *23*, 110–115.

39. WHO (2016). WHO | Grades of hearing impairment. Who.

40. Wallis, C., Ballo, R., Wallis, G., Beighton, P., and Goldblatt, J. (1988). X-linked mixed deafness with stapes fixation in a Mauritian kindred: Linkage to Xq probe pDP34. Genomics *3*, 299–301.

41. HGNC (2019). HUGO Gene Nomenclature - an overview | ScienceDirect Topics.

42. Vona, B., Nanda, I., Hofrichter, M.A.H., Shehata-Dieler, W., and Haaf, T. (2015). Non-syndromic hearing loss gene identification: A brief history and glimpse into the future. Mol. Cell. Probes *29*, 260–270.

43. Snoeckx, R.L., Huygen, P.L.M., Feldmann, D., Marlin, S., Denoyelle, F., Waligora, J., Mueller-Malesinska, M., Pollak, A., Ploski, R., Murgia, A., et al. (2005). GJB2 mutations and degree of hearing loss: A multicenter study. Am. J. Hum. Genet. *77*, 945–957.

44. Yu, Q., Wang, Y., Chang, Q., Wang, J., Gong, S., Li, H., and Lin, X. (2014). Virally expressed connexin26 restores gap junction function in the cochlea of conditional Gjb2 knockout mice. Gene Ther. *21*, 71–80.

45. Wilcox, S.A., Saunders, K., Osborn, A.H., Arnold, A., Wunderlich, J., Kelly, T., Collins, V., Wilcox, L.J., McKinlay Gardner, R.J., Kamarinos, M., et al. (2000). High frequency hearing loss correlated with mutations in the GJB2 gene. Hum. Genet. *106*, 399–405.

46. Wingard, J.C., and Zhao, H.B. (2015). Cellular and deafness mechanisms underlying connexin mutation-induced hearing loss – A common hereditary deafness. Front. Cell. Neurosci. *9*, 202.

47. Zhao, H.B., Yu, N., and Fleming, C.R. (2005). Gap junctional hemichannel-mediated ATP release and hearing controls in the inner ear. Proc. Natl. Acad. Sci. U. S. A. *102*, 18724–18729.

48. Wang, Y., Chang, Q., Tang, W., Sun, Y., Zhou, B., Li, H., and Lin, X. (2009). Targeted connexin26 ablation arrests postnatal development of the organ of Corti. Biochem. Biophys. Res. Commun. *385*, 33–37.

49. Chang, K.W. (2015). Genetics of Hearing Loss-Nonsyndromic. Otolaryngol. Clin. North Am. *48*, 1063–1072.

50. Bykhovskaya, Y., Estivill, X., Taylor, K., Hang, T., Hamon, M., Casano, R.A.M.S., Yang, H., Rotter, J.I., Shohat, M., and Fischel-Ghodsian, N. (2000). Candidate locus for a nuclear modifier gene for maternally iherited deafness. Am. J. Hum. Genet. *66*, 1905–1910.

51. Rodriguez-Paris, J., and Schrijver, I. (2009). The digenic hypothesis unraveled: The GJB6 del(GJB6-D13S1830) mutation causes allele-specific loss of GJB2 expression in cis. Biochem. Biophys. Res. Commun. *389*, 354–359.

52. Liu, X.Z., Yuan, Y., Yan, D., Ding, E.H., Ouyang, X.M., Fei, Y., Tang, W., Yuan, H., Chang, Q., Du, L.L., et al. (2009). Digenic inheritance of non-syndromic deafness caused by mutations at the gap junction proteins Cx26 and Cx31. Hum. Genet. *125*, 53–62.

53. Shahin, H., Walsh, T., Sobe, T., Abu Sa'ed, J., Abu Rayan, A., Lynch, E.D., Lee, M.K., Avraham, K.B., King, M.-C., and Kanaan, M. (2005). Mutations in a Novel Isoform of TRIOBP That Encodes a Filamentous-Actin Binding Protein Are Responsible for DFNB28 Recessive Nonsyndromic Hearing Loss. Am. J. Hum. Genet. *78*, 144–152.

54. Kitajiri, S., Sakamoto, T., Belyantseva, I.A., Goodyear, R.J., Stepanyan, R., Fujiwara, I., Bird, J.E., Riazuddin, S., Riazuddin, S., Ahmed, Z.M., et al. (2010). Actin-Bundling Protein TRIOBP Forms Resilient Rootlets of Hair Cell Stereocilia Essential for Hearing. Cell *141*, 786–798.

55. Boëda, B., El-Amraoui, A., Bahloul, A., Goodyear, R., Daviet, L., Blanchard, S., Perfettini, I., Fath, K.R., Shorte, S., Reiners, J., et al. (2002). Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. EMBO J. *21*, 6689–6699.

56. Liu, X.Z., Walsh, J., Mburu, P., Kendrick-Jones, J., Cope, M.J.T.V., Steel, K.P., and Brown, S.D.M. (1997). Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. Nat. Genet. *16*, 188–190.

57. Puffenberger, E.G., Jinks, R.N., Sougnez, C., Cibulskis, K., Willert, R.A., Achilly, N.P., Cassidy, R.P., Fiorentini, C.J., Heiken, K.F., Lawrence, J.J., et al. (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. PLoS One *7*,.

58. Riazuddin, S., Belyantseva, I.A., Giese, A.P.J., Lee, K., Indzhykulian, A.A., Nandamuri, S.P., Yousaf, R., Sinha, G.P., Lee, S., Terrell, D., et al. (2012). Alterations of the CIB2 calcium-and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48. Nat. Genet. *44*, 1265–1271.

59. Ebermann, I., Phillips, J.B., Liebau, M.C., Koenekoop, R.K., Schermer, B., Lopez, I., Schäfer, E., Roux, A.F., Dafinger, C., Bernd, A., et al. (2010). PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. J. Clin. Invest. *120*, 1812–1823.

60. Mathur, P., and Yang, J. (2015). Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. Biochim. Biophys. Acta - Mol. Basis Dis. *1852*, 406–420.

61. Weil, D., Kussel, P., Blanchard, S., Levy, G., Levi-Acobas, F., Drira, M., Ayadi, H., and Petit, C. (1997). The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. Nat. Genet. *16*, 191–193.

62. Ouyang, X.M., Xia, X.J., Verpy, E., Du, L.L., Pandya, A., Petit, C., Balkany, T., Nance, W.E., and Liu, X.Z. (2002). Mutations in the alternatively spliced exons of USH1C cause non-syndromic recessive deafness. Hum. Genet. *111*, 26–30.

63. Ahmed, Z.M., Smith, T.N., Riazuddin, S., Makishima, T., Ghosh, M., Bokhari, S., Menon, P.S.N., Deshmukh, D., Griffith, A.J., Riazuddin, S., et al. (2002). Nonsyndromic recessive deafness DFNB18 and usher syndrome type IC are allelic mutations of USHIC. Hum. Genet. *110*, 527–531.

64. Bork, J.M., Peters, L.M., Riazuddin, S., Ahmed, Z.M., Li, X.C., Griffith, A.J., Wilcox, E.R., Friedman, T.B., Morell, R.J., Griffith, A.J., et al. (2001). Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23. Am. J. Hum. Genet. *68*, 26–37.

65. Doucette, L., Merner, N.D., Cooke, S., Ives, E., Galutira, D., Walsh, V., Walsh, T., MacLaren, L., Cater, T., Fernandez, B., et al. (2009). Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15. Eur. J. Hum. Genet. *17*, 554–564.

66. Weil, D., El-Amraoui, A., Masmoudi, S., Mustapha, M., Kikkawa, Y., Lainé, S., Delmaghani, S., Adato, A., Nadifi, S., Zina, Z. Ben, et al. (2003). Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. Hum. Mol.

Genet. *12*, 463–471.

67. Weston, M.D., Luijendijk, M.W.J., Humphrey, K.D., Möller, C., and Kimberling, W.J. (2004). Mutations in the VLGR1 Gene Implicate G-Protein Signaling in the Pathogenesis of Usher Syndrome Type II. Am. J. Hum. Genet. *74*, 357–366.

68. Adato, A., Vreugde, S., Joensuu, T., Avidan, N., Hamalainen, R., Belenkiy, O., Olender, T., Bonne-Tamir, B., Ben-Asher, E., Espinos, C., et al. (2002). USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. Eur. J. Hum. Genet. *10*, 339–350.

69. Booth, K.T., Kahrizi, K., Babanejad, M., Daghagh, H., Bademci, G., Arzhangi, S., Zareabdollahi, D., Duman, D., El-Amraoui, A., Tekin, M., et al. (2018). Variants in CIB2 cause DFNB48 and not USH1J. Clin. Genet. *93*, 812–821.

70. Song, J., Feng, Y., Acke, F.R., Coucke, P., Vleminckx, K., and Dhooge, I.J. (2016). Hearing loss in Waardenburg syndrome: A systematic review. Clin. Genet. *89*, 416–425.

71. Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., et al. (1997). Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). Nat. Genet. *17*, 411–422.

72. Sheffield, V.C., Kraiem, Z., Beck, J.C., Nishimura, D., Stone, E.M., Salameh, M., Sadeh, O., and Glaser, B. (1996). Pendred syndrome maps to chromosome 7q21-34 and is caused by an intrinsic defect in thyroid iodine organification. Nat. Genet. *12*, 424–426.

73. Wémeau, J.L., and Kopp, P. (2017). Pendred syndrome. Best Pract. Res. Clin. Endocrinol. Metab. *31*, 213–224.

74. Bizhanova, A., and Kopp, P. (2010). Genetics and phenomics of Pendred syndrome. Mol. Cell. Endocrinol. *322*, 83–90.

75. Lilly, E., Bunick, C.G., Maley, A.M., Zhang, S., Spraker, M.K., Theos, A.J., Vivar, K.L., Seminario-Vidal, L., Bennett, A.E., Sidbury, R., et al. (2019). More than keratitis, ichthyosis, and deafness: Multisystem effects of lethal GJB2 mutations. J. Am. Acad. Dermatol. *80*, 617–625.

76. Fischer, N., Weber, B., and Riechelmann, H. (2016). Presbycusis - Age-related hearing loss. Laryngorhinootologie. *95*, 497–510.

77. Lin, F.R., Thorpe, R., Gordon-Salant, S., and Ferrucci, L. (2011). Hearing loss prevalence and risk factors among older adults in the United States. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. *66 A*, 582–590.

78. Gates, G.A., Cooper, J.C., Kannel, W.B., and Miller, N.J. (1990). Hearing in the elderly: the Framingham cohort, 1983-1985. Part I. Basic audiometric test results. Ear Hear. *11*, 247–256.

79. Helzner, E.P., Cauley, J.A., Pratt, S.R., Wisniewski, S.R., Zmuda, J.M., Talbott, E.O., De Rekeneire, N., Harris, T.B., Rubin, S.M., Simonsick, E.M., et al. (2005). Race and sex differences in age-related hearing loss: The health, aging and body composition study. J. Am. Geriatr. Soc. *53*, 2119–2127.

80. Cruickshanks, K.J., Tweed, T.S., Wiley, T.L., Klein, B.E.K., Klein, R., Chappell, R., Nondahl, D.M., and Dalton, D.S. (2003). The 5-year incidence and progression of hearing loss: The epidemiology of hearing loss study. Arch. Otolaryngol. - Head Neck Surg. *129*, 1041–1046.

81. Roth, T.N., Hanebuth, D., and Probst, R. (2011). Prevalence of age-related hearing loss in Europe: A review. Eur. Arch. Oto-Rhino-Laryngology *268*, 1101–1107.

82. Meyer, C., and Hickson, L. (2012). What factors influence help-seeking for hearing impairment and hearing aid adoption in older adults? Int. J. Audiol. *51*, 66–74.

83. Yamasoba, T., Lin, F.R., Someya, S., Kashio, A., Sakamoto, T., and Kondo, K. (2013). Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways. Hear. Res. *303*, 30–38.

84. Rosenhall, U. (2015). Epidemiology of age related hearing loss. Hear. Balanc. Commun. *13*, 46–50.

85. Agrawal, Y., Platz, E.A., and Niparko, J.K. (2008). Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. Arch. Intern. Med. *168*, 1522–1530.

86. Gopinath, B., Rochtchina, E., Wang, J.J., Schneider, J., Leeder, S.R., and Mitchell, P. (2009). Prevalence of Age-Related Hearing Loss in Older Adults: Blue Mountains Study. Arch. Intern. Med. *169*, 415.

87. Pearson, J.D., Morrell, C.H., Gordon-Salant, S., Brant, L.J., Metter, E.J., Klein, L.L., and Fozard, J.L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. J. Acoust. Soc. Am. *97*, 1196–1205.

88. Smits, C., and Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. Ear Hear. *26*, 89–95.

89. Davis, A.C. (1989). The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int. J. Epidemiol. *18*, 911–917.

90. Lee, J., Dhar, S., Abel, R., Banakis, R., Grolley, E., Lee, J., Zecker, S., and Siegel, J. (2012). Behavioral Hearing Thresholds between 0.125 and 20 kHz Using Depth-Compensated Ear Simulator Calibration. Ear Hear. *33*, 315–329.

91. Gopinath, B., Rochtchina, E., Wang, J.J., Schneider, J., Leeder, S.R., and Mitchell, P. (2009). Prevalence of Age-Related Hearing Loss in Older Adults: Blue Mountains Study. Arch. Intern. Med. *169*, 415.

92. Nolan, L.S., Maier, H., Hermans-Borgmeyer, I., Girotto, G., Ecob, R., Pirastu, N., Cadge, B.A., Hübner, C., Gasparini, P., Strachan, D.P., et al. (2013). Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice. Neurobiol. Aging *34*, 2077.e1-2077.e9.

93. Laer, L. Van, DeStefano, A.L., Myers, R.H., Flothmann, K., Thys, S., Fransen, E., Gates, G.A., Van Camp, G., and Baldwin, C.T. (2002). Is DFNA5 a susceptibility gene for age-related hearing impairment? Eur. J. Hum. Genet. *10*, 883–886.

94. Price, K., Zhu, X., Guimaraes, P.F., Vasilyeva, O.N., and Frisina, R.D. (2009). Hormone replacement therapy diminishes hearing in peri-menopausal mice. Hear. Res. *252*, 29–36.

95. Barrenäs, M.L., and Lindgren, F. (1991). The influence of eye colour on susceptibility to TTS in humans. Br. J. Audiol. *25*, 303–307.

96. Homans, N.C., Metselaar, R.M., Dingemanse, J.G., van der Schroeff, M.P., Brocaar, M.P., Wieringa, M.H., Baatenburg de Jong, R.J., Hofman, A., and Goedegebure, A. (2017). Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. Laryngoscope *127*, 725–730.

97. Agrawal, Y., Platz, E.A., and Niparko, J.K. (2008). Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. Arch. Intern. Med. *168*, 1522–1530.

98. Gates, G.A., Cooper, J.C., Kannel, W.B., and Miller, N.J. (1990). Hearing in the elderly: the Framingham cohort, 1983-1985. Part I. Basic audiometric test results. Ear Hear. *11*, 247–256.

99. Murillo-Cuesta, S., Contreras, J., Zurita, E., Cediel, R., Cantero, M., Varela-Nieto, I., and Montoliu, L. (2010). Melanin precursors prevent premature age-related and noise-induced hearing loss in albino mice. Pigment Cell Melanoma Res. *23*, 72–83.

100. Bedin, E., Franzè, A., Zadro, C., Persico, M.G., Ciullo, M., Hladnik, U., Dolcetta, D., Grasso, D.L., Riccardi, P., Nutile, T., et al. (2009). Age-related hearing loss in four Italian genetic isolates: An epidemiological study. Int. J. Audiol. *48*, 465–472.

101. Momi, S.K., Wolber, L.E., Fabiane, S.M., MacGregor, A.J., and Williams, F.M.K. (2015). Genetic and Environmental Factors in Age-Related Hearing Impairment. Twin Res. Hum. Genet. *18*, 383–392.

102. Wolber, L.E., Steves, C.J., Spector, T.D., and Williams, F.M.K. (2012). Hearing ability with age in Northern European women: A new web-based approach to genetic studies. PLoS One *7*, e35500.

103. Bogo, R., Farah, A., Johnson, A.C., Karlsson, K.K., Pedersen, N.L., Svartengren, M., and Skjönsberg, Å. (2015). The Role of Genetic Factors for Hearing Deterioration Across 20 Years: A Twin Study. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. *70*, 647–653.

104. Viljanen, A., Kaprio, J., Pyykkö, I., Sorri, M., Kauppinen, M., Koskenvuo, M., and Rantanen, T. (2007). Genetic and environmental influences on hearing at different frequencies separately for the

better and worse hearing ear in older women. Int. J. Audiol. *46*, 772–779.

105. Hendrickx, J.J., Huyghe, J.R., Topsakal, V., Demeester, K., Wienker, T.F., Van Laer, L., Van Eyken, E., Fransen, E., Mäki-Torkko, E., Hannula, S., et al. (2013). Familial aggregation of pure tone hearing thresholds in an aging European population. Otol. Neurotol. *34*, 838–844.

106. Newman, D.L., Fisher, L.M., Ohmen, J., Parody, R., Fong, C.T., Frisina, S.T., Mapes, F., Eddins, D.A., Frisina, R.D., Frisina, R.D., et al. (2012). GRM7 variants associated with age-related hearing loss based on auditory perception. Hear. Res. *294*, 125–132.

107. Van Laer, L., Huyghe, J.R., Hannula, S., Van Eyken, E., Stephan, D.A., Mäki-Torkko, E., Aikio, P., Fransen, E., Lysholm-Bernacchi, A., Sorri, M., et al. (2010). A genome-wide association study for age-related hearing impairment in the Saami. Eur. J. Hum. Genet. *18*, 685–693.

108. Luo, H., Yang, T., Jin, X., Pang, X., Li, J., Chai, Y., Li, L., Zhang, Y., Zhang, L., Zhang, Z., et al. (2013). Association of GRM7 Variants with Different Phenotype Patterns of Age-Related Hearing Impairment in an Elderly Male Han Chinese Population. PLoS One *8*, e77153.

109. Friedman, R.A., Van Laer, L., Huentelman, M.J., Sheth, S.S., Van Eyken, E., Corneveaux, J.J., Tembe, W.D., Halperin, R.F., Thorburn, A.Q., Thys, S., et al. (2009). GRM7 variants confer susceptibility to age-related hearing impairment. Hum. Mol. Genet. *18*, 785–796.

110. Duijvestijn, J.A., Anteunis, L.J., Hendriks, J.J., and Manni, J.J. (1999). Definition of hearing impairment and its effect on prevalence figures. A survey among senior citizens. Acta Otolaryngol. *119*, 420–423.

111. Wolber, L.E., Girotto, G., Buniello, A., Vuckovic, D., Pirastu, N., Lorente-Cánovas, B., Rudan, I., Hayward, C., Polasek, O., Ciullo, M., et al. (2014). Salt-inducible kinase 3, SIK3, is a new gene associated with hearing. Hum. Mol. Genet. *23*, 6407–6418.

112. Vuckovic, D., Dawson, S., Scheffer, D.I., Rantanen, T., Morgan, A., Di Stazio, M., Vozzi, D., Nutile, T., Concas, M.P., Biino, G., et al. (2015). Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. Hum. Mol. Genet. *24*, 5655–5664.

113. Hoffmann, T.J., Keats, B.J., Yoshikawa, N., Schaefer, C., Risch, N., and Lustig, L.R. (2016). A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. PLoS Genet. *12*, e1006371.

114. Girotto, G., Pirastu, N., Sorice, R., Biino, G., Campbell, H., d'Adamo, A.P., Hastie, N.D., Nutile, T., Polasek, O., Portas, L., et al. (2011). Hearing function and thresholds: A genome-wide association study in European isolated populations identifies new loci and pathways. J. Med. Genet. *48*, 369–374.

115. Nolte, I.M., Jansweijer, J.A., Riese, H., Asselbergs, F.W., Van Der Harst, P., Spector, T.D., Pinto, Y.M., Snieder, H., and Jamshidi, Y. (2017). A Comparison of Heritability Estimates by Classical Twin

Modeling and Based on Genome-Wide Genetic Relatedness for Cardiac Conduction Traits. Twin Res. Hum. Genet. *20*, 489–498.

116. Young, A.I. (2019). Solving the missing heritability problem. PLoS Genet. *15*, e1008222.

117. Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. Soc. Sci. Res. *43*, 184–199.

118. Kvestad, E., Czajkowski, N., Krog, N.H., Engdahl, B., and Tambs, K. (2012). Heritability of hearing loss. Epidemiology *23*, 328–331.

119. Christensen, K., Frederiksen, H., and Hoffman, H.J. (2001). Genetic and environmental influences on self-reported reduced hearing in the old and oldest old. J. Am. Geriatr. Soc. *49*, 1512–1517.

120. Van Eyken, E., Van Camp, G., and Van Laer, L. (2007). The complexity of age-related hearing impairment: Contributing environmental and genetic factors. Audiol. Neurotol. *12*, 345–358.

121. Rigters, S.C., Metselaar, M., Wieringa, M.H., De Jong, R.J.B., Hofman, A., and Goedegebure, A. (2016). Contributing Determinants to Hearing Loss in Elderly Men and Women: Results from the Population-Based Rotterdam Study. Audiol. Neurotol. *21*, 10–15.

122. Moore, D.R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R.H., and Munro, K.J. (2014). Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age. PLoS One *9*, e107720.

123. Dawes, P., Fortnum, H., Moore, D.R., Emsley, R., Norman, P., Cruickshanks, K., Davis, A., Edmondson-Jones, M., McCormack, A., Lutman, M., et al. (2014). Hearing in middle age: A population snapshot of 40- to 69-year olds in the united kingdom. Ear Hear. *35*, e44–e51.

124. Cruickshanks, K.J., Nondahl, D.M., Dalton, D.S., Fischer, M.E., Klein, B.E.K., Klein, R., Nieto, F.J., Schubert, C.R., and Tweed, T.S. (2015). Smoking, central adiposity, and poor glycemic control increase risk of hearing impairment. J. Am. Geriatr. Soc. *63*, 918–924.

125. Ryan, A.F., Kujawa, S.G., Hammill, T., Le Prell, C., and Kil, J. (2016). Temporary and Permanent Noise-induced Threshold Shifts: A Review of Basic and Clinical Observations. Otol. Neurotol. *37*, e271–e275.

126. Nondahl, D.M., Shi, X., Cruickshanks, K.J., Dalton, D.S., Tweed, T.S., Wiley, T.L., and Carmichael, L.L. (2009). Notched audiograms and noise exposure history in older adults. Ear Hear. *30*, 696–703.

127. Gates, G.A., Schmid, P., Kujawa, S.G., Nam, B.H., and D'Agostino, R. (2000). Longitudinal threshold changes in older men with audiometric notches. Hear. Res. *141*, 220–228.

128. Kujawa, S.G., and Liberman, M.C. (2006). Acceleration of age-related hearing loss by early noise exposure: Evidence of a misspent youth. J. Neurosci. *26*, 2115–2123.

129. Śliwińska-Kowalska, M., and Zaborowski, K. (2017). WHO environmental noise guidelines for

the European region: A systematic review on environmental noise and permanent hearing loss and tinnitus. Int. J. Environ. Res. Public Health *14*, 1139.

130. Sheth, S., Mukherjea, D., Rybak, L.P., and Ramkumar, V. (2017). Mechanisms of Cisplatin-Induced Ototoxicity and Otoprotection. Front. Cell. Neurosci. *11*, 338.

131. Dulon, D., Hiel, H., Aurousseau, C., Erre, J.P., and Aran, J.M. (1993). Pharmacokinetics of gentamicin in the sensory hair cells of the organ of Corti: Rapid uptake and long term persistence. Comptes Rendus l'Academie Des Sci. - Ser. III *316*, 682–687.

132. Aran, J.M., Erre, J.P., Lima Da Costa, D., Debbarh, I., and Dulon, D. (1999). Acute and chronic effects of aminoglycosides on cochlear hair cells. Ann. N. Y. Acad. Sci. *884*, 60–68.

133. Campo, P., Morata, T.C., and Hong, O.S. (2013). Chemical exposure and hearing loss. Disease-a-Month *59*, 119–138.

134. Campo P, Ma guin K, Gabriel S, et al. (2009). Combined exposure to noise and ototoxic substances: European Agency for Safety and Health at Work (EU-OSHA).

135. Sugiura, S., Uchida, Y., Nakashima, T., Ando, F., and Shimokata, H. (2010). The association between gene polymorphisms in uncoupling proteins and hearing impairment in Japanese elderly. Acta Otolaryngol. *130*, 487–492.

136. Nolan, L.S., Cadge, B.A., Gomez-Dorado, M., and Dawson, S.J. (2013). A functional and genetic analysis of SOD2 promoter variants and their contribution to age-related hearing loss. Mech. Ageing Dev. *134*, 298–306.

137. Seidman, M.D. (2000). Effects of dietary restriction and antioxidants on presbyacusis. Laryngoscope *110*, 727–738.

138. Henry, K.R. (1986). Effects of dietary restriction on presbyacusis in the mouse1. Int. J. Audiol. *25*, 329–337.

139. Someya, S., Yamasoba, T., Weindruch, R., Prolla, T.A., and Tanokura, M. (2007). Caloric restriction suppresses apoptotic cell death in the mammalian cochlea and leads to prevention of presbycusis. Neurobiol. Aging *28*, 1613–1622.

140. Cruickshanks, K.J., Klein, R., Klein, B.E.K., Wiley, T.L., Nondahl, D.M., and Tweed, T.S. (1998). Cigarette smoking and hearing loss: The epidemiology of hearing loss study. J. Am. Med. Assoc. *279*, 1715–1719.

141. Dawes, P., Cruickshanks, K.J., Moore, D.R., Edmondson-Jones, M., McCormack, A., Fortnum, H., and Munro, K.J. (2014). Cigarette smoking, passive smoking, alcohol consumption, and hearing loss. JARO - J. Assoc. Res. Otolaryngol. *15*, 663–674.

142. Harkrider, A.W., Champlin, C.A., and McFadden, D. (2001). Acute effect of nicotine on non-

smokers: I. OAEs and ABRs. Hear. Res. *160*, 73–88.

143. Nakanishi, N., Okamoto, M., Nakamura, K., Suzuki, K., and Tatara, K. (2000). Cigarette smoking and risk for hearing impairment: A longitudinal study in Japanese male office workers. J. Occup. Environ. Med. *42*, 1045–1049.

144. Torre, P., Cruickshanks, K.J., Klein, B.E.K., Klein, R., and Nondahl, D.M. (2006). The Association Between Cardiovascular Disease and Cochlear Function in Older Adults. J. Speech, Lang. Hear. Res. *48*, 473–481.

145. Brant, L.J., Gordon-Salant, S., Pearson, J.D., Klein, L.L., Morrell, C.H., Metter, E.J., and Fozard, J.L. (1996). Risk factors related to age-associated hearing loss in the speech frequencies. J. Am. Acad. Audiol. *7*, 152–160.

146. Gates, G.A., Cobb, J.L., D'Agostino, R.B., and Wolf, P.A. (1993). The relation of hearing in the elderly to the presence of cardiovascular disease and cardiovascular risk factors. Arch. Otolaryngol. Head. Neck Surg. *119*, 156–161.

147. Mitchell, P., Gopinath, B., McMahon, C.M., Rochtchina, E., Wang, J.J., Boyages, S.C., and Leeder, S.R. (2009). Relationship of Type 2 diabetes to the prevalence, incidence and progression of age-related hearing loss. Diabet. Med. *26*, 483–488.

148. Kiely, K.M., Gopinath, B., Mitchell, P., Luszcz, M., and Anstey, K.J. (2012). Cognitive, health, and sociodemographic predictors of longitudinal decline in hearing acuity among older adults. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. *67 A*, 997–1003.

149. Rönnberg, J., Hygge, S., Keidser, G., and Rudner, M. (2014). The effect of functional hearing loss and age on long- and short-term visuospatial memory: Evidence from the UK biobank resource. Front. Aging Neurosci. *6*, 1–13.

150. Rönnberg, J., Rudner, M., and Lunner, T. (2011). Cognitive Hearing Science: The Legacy of Stuart Gatehouse. Trends Amplif. *15*, 140–148.

151. Tay, T., Jie, J.W., Kifley, A., Lindley, R., Newall, P., and Mitchell, P. (2006). Sensory and cognitive association in older persons: Findings from an older Australian population. Gerontology *52*, 386–394.

152. Uhlmann, R.F., Larson, E.B., Rees, T.S., Koepsell, T.D., and Duckert, L.G. (1989). Relationship of hearing impairment to dementia and cognitive dysfunction in older adults. JAMA *261*, 1916–1919.

153. Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S.G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., et al. (2017). Dementia prevention, intervention, and care. Lancet *390*, 2673–2734.

154. Fulton, S.E., Lister, J.J., Bush, A.L.H., Edwards, J.D., and Andel, R. (2015). Mechanisms of the Hearing-Cognition Relationship. Semin. Hear. *36*, 140–149.

155. Rabbitt, P.M. (1968). Channel-capacity, intelligibility and immediate memory. Q. J. Exp. Psychol. *20*, 241–248.

156. Rabbitt, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. Acta Otolaryngol. *111*, 167–176.

157. McCoy, S.L., Tun, P.A., Cox, L.C., Colangelo, M., Stewart, R.A., and Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol. *58*, 22–33.

158. Tun, P.A., O'Kane, G., and Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. Psychol. Aging *17*, 453–467.

159. Boxtel, M.P.J. Van, Beijsterveldt, C.E.M. Van, Houx, P.J., and Anteunis, L.J.C. (2000). Mild Hearing Impairment Can Reduce Verbal Memory Per- formance in a Healthy Adult Population. J. Clin. Exp. Neuropsychol. *22*, 147–154.

160. Rönnberg, J., Hygge, S., Keidser, G., and Rudner, M. (2014). The effect of functional hearing loss and age on long- and short-term visuospatial memory: Evidence from the UK biobank resource. Front. Aging Neurosci. *6*, 326.

161. Heinrich, A., Henshaw, H., and Ferguson, M.A. (2016). Only behavioral but not self-report measures of speech perception correlate with cognitive abilities. Front. Psychol. *7*, 1–16.

162. Dawes, P., Emsley, R., Cruickshanks, K.J., Moore, D.R., Fortnum, H., Edmondson-Jones, M., McCormack, A., and Munro, K.J. (2015). Hearing loss and cognition: The role of hearing aids, social isolation and depression. PLoS One *10*, e0119616.

163. Young Choi, A., Shim, H.J., Lee, S.H., Yoon, S.W., and Joo, E.J. (2011). Is cognitive function in adults with hearing impairment improved by the use of hearing aids? Clin. Exp. Otorhinolaryngol. *4*, 72–76.

164. Acar, B., Yurekli, M.F., Babademez, M.A., Karabulut, H., and Karasen, R.M. (2011). Effects of hearing aids on cognitive functions and depressive signs in elderly people. Arch. Gerontol. Geriatr. *52*, 250–252.

165. Nondahl, D.M., Cruickshanks, K.J., Huang, G.H., Klein, B.E.K., Klein, R., Javier Nieto, F., and Tweed, T.S. (2011). Tinnitus and its risk factors in the Beaver Dam Offspring Study. Int. J. Audiol. *50*, 313–320.

166. Nondahl, D.M., Cruickshanks, K.J., Wiley, T.L., Klein, B.E.K., Klein, R., Chappell, R., and Tweed, T.S. (2010). The ten-year incidence of tinnitus among older adults. Int. J. Audiol. *49*, 580–585.

167. Bhatt, J.M., Lin, H.W., and Bhattacharyya, N. (2016). Tinnitus Epidemiology: Prevalence, Severity, Exposures And Treatment Patterns In The United States: Bhatt JM: Tinnitus in the United States HHS Public Access. JAMA Otolaryngol Head Neck Surg *142*, 959–965.

168. Crocetti, A., Forti, S., Ambrosetti, U., and Bo, L. Del (2009). Questionnaires to evaluate anxiety and depressive levels in tinnitus patients. Otolaryngol. - Head Neck Surg. *140*, 403–405.

169. Krog, N.H., Engdahl, B., and Tambs, K. (2010). The association between tinnitus and mental health in a general population sample: Results from the HUNT Study. J. Psychosom. Res. *69*, 289–298.

170. Sand, P.G., Langguth, B., Kleinjung, T., and Eichhammer, P. (2007). Genetics of chronic tinnitus. Prog. Brain Res. *166*, 159–168.

171. Schecklmann, M., Vielsmeier, V., Steffens, T., Landgrebe, M., Langguth, B., and Kleinjung, T. (2012). Relationship between audiometric slope and tinnitus pitch in tinnitus patients: Insights into the mechanisms of tinnitus generation. PLoS One *7*, e34878.

172. Norena, A., Micheyl, C., Chéry-Croze, S., and Collet, L. (2002). Psychoacoustic characterization of the tinnitus spectrum: Implications for the underlying mechanisms of tinnitus. Audiol. Neuro-Otology *7*, 358–369.

173. Krauss, P., Schilling, A., Tziridis, K., and Schulze, H. (2019). Models of tinnitus development: From cochlea to cortex. HNO *67*, 172–177.

174. Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., and Schulze, H. (2016). Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front. Neurosci. *10*, 597.

175. Soussi, T., and Otto, S.R. (1994). Effects of electrical brainstem stimulation on tinnitus. Acta Otolaryngol. *114*, 135–140.

176. Kvestad, E., Czajkowski, N., Engdahl, B., Hoffman, H.J., and Tambs, K. (2010). Low heritability of tinnitus: Results from the second Nord-Trøndelag Health Study. Arch. Otolaryngol. - Head Neck Surg. *136*, 178–182.

177. Maas, I.L., Brüggemann, P., Requena, T., Bulla, J., Edvall, N.K., Hjelmborg, J.V.B., Szczepek, A.J., Canlon, B., Mazurek, B., Lopez-Escamez, J.A., et al. (2017). Genetic susceptibility to bilateral tinnitus in a Swedish twin cohort. Genet. Med. *19*, 1007–1012.

178. Bogo, R., Farah, A., Karlsson, K.K., Pedersen, N.L., Svartengren, M., and Skjönsberg, Å. (2016). Prevalence, Incidence Proportion, and Heritability for Tinnitus. Ear Hear. *38*, 292–300.

179. Jayakody, D.M.P., Almeida, O.P., Speelman, C.P., Bennett, R.J., Moyle, T.C., Yiannos, J.M., and Friedland, P.L. (2018). Association between speech and high-frequency hearing loss and depression, anxiety and stress in older adults. Maturitas *110*, 86–91.

180. Lee, A.T.H., Tong, M.C.F., Yuen, K.C.P., Tang, P.S.O., and Van Hasselt, C.A. (2010). Hearing impairment and depressive symptoms in an older chinese population. J. Otolaryngol. - Head Neck Surg. *39*, 498–503.

181. Gopinath, B., Wang, J.J., Schneider, J., Burlutsky, G., Snowdon, J., McMahon, C.M., Leeder, S.R., and Mitchell, P. (2009). Depressive symptoms in older adults with hearing impairments: The blue mountains study: Letters to the editor. J. Am. Geriatr. Soc. *57*, 1306–1308.

182. Cosh, S., von Hanno, T., Helmer, C., Bertelsen, G., Delcourt, C., and Schirmer, H. (2018). The association amongst visual, hearing, and dual sensory loss with depression and anxiety over 6 years: The Tromsø Study. Int. J. Geriatr. Psychiatry *33*, 598–605.

183. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

184. Smith, G.D. (2010). Mendelian randomization for strengthening causal inference in observational studies: Application to gene × environment interactions. Perspect. Psychol. Sci. *5*, 527–545.

185. Schaette, R., and McAlpine, D. (2011). Tinnitus with a Normal Audiogram: Physiological Evidence for Hidden Hearing Loss and Computational Model. J. Neurosci. *31*, 13452–13457.

186. Smits, C., Theo Goverts, S., and Festen, J.M. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. J. Acoust. Soc. Am. *133*, 1693–1706.

187. Plomp, R. (1976). Binaural and Monaural Speech Intelligibility of Connected Discourse in Reverberation as a Function of Azimuth of a Single Competing Sound Source (Speech or Noise). Acta Acust. United with Acust. *34*, 200–211.

188. Smits, C., Kapteyn, T.S., and Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. Int. J. Audiol. *43*, 15–28.

189. Jansen, S., Luts, H., Dejonckere, P., Van Wieringen, A., and Wouters, J. (2013). Efficient hearing screening in noise-exposed listeners using the digit triplet test. Ear Hear. *34*, 773–778.

190. Smits, C., Theo Goverts, S., and Festen, J.M. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. J. Acoust. Soc. Am. *133*, 1693–1706.

191. Zokoll, M.A., Wagener, K.C., Brand, T., Buschermöhle, M., and Kollmeier, B. (2012). Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype. Int. J. Audiol. *51*, 697–707.

192. Ozimek, E., Kutzner, D., Sk, A., and Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. Int. J. Audiol. *48*, 433–443.

193. Jansen, S., Luts, H., Wagener, K.C., Frachet, B., and Wouters, J. (2010). The French digit triplet test: A hearing screening tool for speech intelligibility in noise. Int. J. Audiol. *49*, 378–387.

194. Hall, S. (2006). The Development of a New English Sentence in Noise Test and an English Number

Recognition Test. (Unpublished MSc Thesis).

195. Keidser, G., Seeto, M., Rudner, M., Hygge, S., and Rönnberg, J. (2015). On the relationship between functional hearing and depression. Int. J. Audiol. *54*, 653–664.

196. Martini, A., Mazzoli, M., Rosignoli, M., Trevisi, P., Maggi, S., Enzi, G., and Crepaldi, G. (2008). Hearing in the Elderly: A Population Study: Audición en la senectud: Un estudio poblacional. Int. J. Audiol. *40*, 285–291.

197. Davis, A., Smith, P., Ferguson, M., and Stephens, D. (2007). Acceptability, benefit and costs of early screening for hearing disability. Health Technol. Assess. (Rockv). *11*, 1–293.

198. Davis, A.C. (1989). The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int. J. Epidemiol. *18*, 911–917.

199. Hietanen, A., Era, P., Sorri, M., and Heikkinen, E. (2004). Changes in hearing in 80-year-old people: a 10-year follow-up study. Int. J. Audiol. *43*, 126–135.

200. Hietanen, A., Era, P., Henrichsen, J., Rosenhall, U., Sorri, M., and Heikkinen, E. (2005). Hearing among 75-year-old people in three Nordic localities: a comparative study. Int. J. Audiol. *44*, 500–508.

201. Rosenhall, U., Jönsson, R., and Söderlind, O. Self-assessed hearing problems in Sweden: a demographic study. Audiology *38*, 328–334.

202. Choi, J.S., Betz, J., Deal, J., Contrera, K.J., Genther, D.J., Chen, D.S., Gispen, F.E., and Lin, F.R. (2016). A Comparison of Self-Report and Audiometric Measures of Hearing and Their Associations with Functional Outcomes in Older Adults. J. Aging Health *28*, 890–910.

203. Girotto, G., Vuckovic, D., Buniello, A., Lorente-Cánovas, B., Lewis, M., Gasparini, P., and Steel, K.P. (2014). Expression and replication studies to identify new candidate genes involved in normal hearing function. PLoS One *9*, e85352.

204. Makishima, T., Madeo, A.C., Brewer, C.C., Zalewski, C.K., Butman, J.A., Sachdev, V., Arai, A.E., Holbrook, B.M., Rosing, D.R., and Griffith, A.J. (2007). Nonsyndromic hearing loss DFNA10 and a novel mutation of EYA4: Evidence for correlation of normal cardiac phenotype with truncating mutations of the Eya domain. Am. J. Med. Genet. Part A *143*, 1592–1598.

205. Yu, C., Meng, X., Zhang, S., Zhao, G., Hu, L., and Kong, X. (2003). A 3-nucleotide deletion in the polypyrimidine tract of intron 7 of the DFNA5 gene causes nonsyndromic hearing impairment in a Chinese family. Genomics *82*, 575–579.

206. DeStefano, A.L., Gates, G.A., Heard-Costa, N., Myers, R.H., and Baldwin, C.T. (2003). Genomewide linkage analysis to presbycusis in the Framingham Heart Study. Arch. Otolaryngol. Head. Neck Surg. *129*, 285–289.

207. Garringer, H.J., Pankratz, N.D., Nichols, W.C., and Reed, T. (2006). Hearing impairment

susceptibility in elderly men and the DFNA18 locus. Arch. Otolaryngol. - Head Neck Surg. *132*, 506–510.

208. Huyghe, J.R., Van Laer, L., Hendrickx, J.J., Fransen, E., Demeester, K., Topsakal, V., Kunst, S., Manninen, M., Jensen, M., Bonaconsa, A., et al. (2008). Genome-wide SNP-Based Linkage Scan Identifies a Locus on 8q24 for an Age-Related Hearing Impairment Trait. Am. J. Hum. Genet. *83*, 401–407.

209. Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. *6*, 95–108.

210. Patnala, R., Clements, J., and Batra, J. (2013). Candidate gene association studies: A comprehensive guide to useful in silico tools.

211. Van Eyken, E., Van Camp, G., Fransen, E., Topsakal, V., Hendrickx, J.J., Demeester, K., Van De Heyning, P., Mäki-Torkko, E., Hannula, S., Sorri, M., et al. (2007). Contribution of the N-acetyltransferase 2 polymorphism NAT2*6A to age-related hearing impairment. J. Med. Genet. *44*, 570–578.

212. Van Laer, L., Van Eyken, E., Fransen, E., Huyghe, J.R., Topsakal, V., Hendrickx, J.J., Hannula, S., Mäki-Torkko, E., Jensen, M., Demeester, K., et al. (2008). The grainyhead like 2 gene (GRHL2), alias TFCP2L3, is associated with age-related hearing impairment. Hum. Mol. Genet. *17*, 159–169.

213. Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-Wide Association Studies. PLoS Comput. Biol. *8*, e1002822.

214. Gates, G.A., and Mills, J.H. (2005). Presbycusis. Lancet *366*, 1111–1120.

215. Chang, N.-C., Dai, C.-Y., Lin, W.-Y., Yang, H.-L., Wang, H.-M., Chien, C.-Y., Hsieh, M.-H., and Ho, K.-Y. (2018). The Association of GRM7 Single Nucleotide Polymorphisms with Age-Related Hearing Impairment in a Taiwanese Population. J. Int. Adv. Otol. *14*, 170–175.

216. Yu, P., Jiao, J., Chen, G., Zhou, W., Zhang, H., Wu, H., Li, Y., Gu, G., Zheng, Y., Yu, Y., et al. (2018). Effect of GRM7 polymorphisms on the development of noise-induced hearing loss in Chinese Han workers: A nested case-control study. BMC Med. Genet. *19*, 4.

217. Haider, H.F., Flook, M., Aparicio, M., Ribeiro, D., Antunes, M., Szczepek, A.J., Hoare, D.J., Fialho, G., Paço, J.C., and Caria, H. (2017). Biomarkers of presbycusis and tinnitus in a Portuguese older population. Front. Aging Neurosci. *9*, 346.

218. Richards, J.B., Kavvoura, F.K., Rivadeneira, F., Styrkársdóttir, U., Estrada, K., Halldórsson, B. V., Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Mullin, B.H., et al. (2009). Collaborative meta-analysis: Associations of 150 candidate genes with osteoporosis and osteoporotic fracture. Ann. Intern. Med. *151*, 528–537.

219. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017).

10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22.

220. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. *109*, 1193–1198.

221. Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature *465*, 721–727.

222. Wolber, L.E., Steves, C.J., Tsai, P.C., Deloukas, P., Spector, T.D., Bell, J.T., and Williams, F.M.K. (2014). Epigenome-wide DNA methylation in hearing ability: New mechanisms for an old problem. PLoS One *9*, e105729.

223. Gao, X.R., Huang, H., and Kim, H. (2019). Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. Hum. Mol. Genet. *28*, 1162–1172.

224. Khawaja, A.P., Cooke Bailey, J.N., Wareham, N.J., Scott, R.A., Simcoe, M., Igo, R.P., Song, Y.E., Wojciechowski, R., Cheng, C.Y., Khaw, P.T., et al. (2018). Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. Nat. Genet. *50*, 778–782.

225. Marioni, R.E., Harris, S.E., Zhang, Q., McRae, A.F., Hagenaars, S.P., Hill, W.D., Davies, G., Ritchie, C.W., Gale, C.R., Starr, J.M., et al. (2018). GWAS on family history of Alzheimer's disease. Transl. Psychiatry *8*, 99.

226. Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. *47*, 284–290.

227. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

228. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A., and Price, A. (2017). Mixed model association for biobank-scale data sets. BioRxiv 194944.

229. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.

230. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. *39*, 906–913.

231. Kraft, P., Zeggini, E., and Ioannidis, J.P.A. (2009). Replication in genome-wide association studies. Stat. Sci. *24*, 561–573.

232. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G.,

Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375.

233. Watanabe, K., Taskesen, E., Van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.

234. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput. Biol. *11*, e1004219.

235. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 1–9.

236. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

237. Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganella, S., Min, J.L., Walter, K., Timpson, N.J., Dunham, I., Birney, E., and Soranzo, N. (2019). GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat. Genet. *51*, 343–353.

238. Bowl, M.R., and Dawson, S.J. (2015). The mouse as a model for age-related hearing loss - A mini-review. Gerontology *61*, 149–157.

239. Brown, S.D.M., Hardisty-Hughes, R.E., and Mburu, P. (2008). Quiet as a mouse: Dissecting the molecular and genetic basis of hearing. Nat. Rev. Genet. *9*, 277–290.

240. György, B., Nist-Lund, C., Pan, B., Asai, Y., Karavitaki, K.D., Kleinstiver, B.P., Garcia, S.P., Zaborowski, M.P., Solanes, P., Spataro, S., et al. (2019). Allele-specific gene editing prevents deafness in a model of dominant progressive hearing loss. Nat. Med. *25*, 1123–1130.

241. Kane, K.L., Longo-Guess, C.M., Gagnon, L.H., Ding, D., Salvi, R.J., and Johnson, K.R. (2012). Genetic background effects on age-related hearing loss associated with Cdh23 variants in mice. Hear. Res. *283*, 80–88.

242. Crow, A.L., Ohmen, J., Wang, J., Lavinsky, J., Hartiala, J., Li, Q., Li, X., Salehide, P., Eskin, E., Pan, C., et al. (2015). The genetic architecture of hearing impairment in mice: Evidence for frequency-specific genetic determinants. G3 Genes, Genomes, Genet. *5*, 2329–2339.

243. Bowl, M.R., Simon, M.M., Ingham, N.J., Greenaway, S., Santos, L., Cater, H., Taylor, S., Mason, J., Kurbatova, N., Pearson, S., et al. (2017). A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. Nat. Commun. *8*, 886.

244. Ingham, N.J., Carlisle, F., Pearson, S., Lewis, M.A., Buniello, A., Chen, J., Isaacson, R.L., Pass, J., White, J.K., Dawson, S.J., et al. (2016). S1PR2 variants associated with auditory function in humans and endocochlear potential decline in mouse. Sci. Rep. *6*, 1–13.

245. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. *12*, e1001779.

246. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants Supplementary Material. BioRxiv 166298.

247. Marchini, J., O'Connell, J., Delaneau, O., Sharp, K., Kretzschmar, W., Band, G., McCarthy, S., Petkova, D., Bycroft, C., Freeman, C., et al. Genotype imputation and genetic association studies of UK Biobank. PDF access via: www.ukbiobank.ac.uk/scient.

248. Moayyeri, A., Hammond, C.J., Hart, D.J., and Spector, T.D. (2013). The UK adult twin registry (twinsUK resource). Twin Res. Hum. Genet. *16*, 144–149.

249. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

250. Steptoe, A., Breeze, E., Banks, J., and Nazroo, J. (2013). Cohort profile: The English Longitudinal Study of Ageing. Int. J. Epidemiol. *42*, 1640–1648.

251. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., et al. (2017). LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272–279.

252. Bulik-Sullivan, B., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., Corvin, A., et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

253. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

254. De Leeuw, C.A., Neale, B.M., Heskes, T., and Posthuma, D. (2016). The statistical properties of gene-set analysis. Nat. Rev. Genet. *17*, 353–364.

255. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

256. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., and Frishman, D. (2011). LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics *27*, 2336–2337.

257. Machiela, M.J., and Chanock, S.J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.

Bioinformatics *31*, 3555–3557.

258. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

259. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885.

260. Lewis, M.A., Nolan, L.S., Cadge, B.A., Matthews, L.J., Schulte, B.A., Dubno, J.R., Steel, K.P., and Dawson, S.J. (2018). Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. BMC Med. Genomics *11*, 77.

261. Cai, T., Jen, H.I., Kang, H., Klisch, T.J., Zoghbi, H.Y., and Groves, A.K. (2015). Characterization of the transcriptome of nascent hair cells and identification of direct targets of the atoh1 transcription factor. J. Neurosci. *35*, 5870–5883.

262. Spongr, V.P., Flood, D.G., Frisina, R.D., and Salvi, R.J. (1997). Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans. J. Acoust. Soc. Am. *101*, 3546–3553.

263. NHS (2019). What is the body mass index (BMI)? - NHS.

264. Coggon, D., Barker, D., and Rose, G. (2009). Chapter 2. Quantifying disease in populations. | The BMJ.

265. Coggen, D., Rose, G., & Barker, D.J.P. (2012). Chapter 8. Case-control and cross sectional studies | The BMJ.

266. Fischer, N., Weber, B., and Riechelmann, H. (2016). Presbyakusis. Laryngorhinootologie. *95*, 497–510.

267. Keidser, G., Rudner, M., Seeto, M., Hygge, S., and Rönnberg, J. (2016). The effect of functional hearing and hearing aid usage on verbal reasoning in a large community-dwelling population. Ear Hear. *37*, e26–e36.

268. Dawes, P., Cruickshanks, K.J., Moore, D.R., Fortnum, H., Edmondson-Jones, M., McCormack, A., and Munro, K.J. (2015). The effect of prenatal and childhood development on hearing, vision and cognition in adulthood. PLoS One *10*, e0136590.

269. Vlaming, M.S.M.G., Mackinnon, R.C., Jansen, M., and Moore, D.R. (2014). Automated screening for high-frequency hearing loss. Ear Hear. *35*, 667–679.

270. Leensen, M.C.J., De Laat, J.A.P.M., Snik, A.F.M., and Dreschler, W.A. (2011). Speech-in-noise screening tests by internet, Part 2: Improving test sensitivity for noise-induced hearing loss. Int. J. Audiol. *50*, 835–848.

271. Dawes, P. (2013). UK Biobank normative data for the Digit Triplet Test. Access via www.ukbiobank.ac.uk. 1–4.

272. Stam, M., Smits, C., Twisk, J.W.R., Lemke, U., Festen, J.M., and Kramer, S.E. (2015). Deterioration of Speech Recognition Ability Over a Period of 5 Years in Adults Ages 18 to 70 Years. Ear Hear. *36*, e129–e137.

273. Smits, C., Kramer, S.E., and Houtgast, T. (2006). Speech reception thresholds in noise and self-reported hearing disability in a general adult population. Ear Hear. *27*, 538–549.

274. Goh, L., and Yap, V.B. (2009). Effects of normalization on quantitative traits in association test. BMC Bioinformatics *10*, 415.

275. Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N.E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. Am. J. Epidemiol. *186*, 1026–1034.

276. Smits, C. (2017). Improving the efficiency of speech-in-noise hearing screening tests. Ear Hear. *38*, e385–e388.

277. Momi, S.K., Wolber, L.E., Fabiane, S.M., MacGregor, A.J., and Williams, F.M.K. (2015). Genetic and Environmental Factors in Age-Related Hearing Impairment. Twin Res. Hum. Genet. *18*, 383–392.

278. Kim, S.Y., Kim, H.J., Kim, M.S., Park, B., Kim, J.H., and Choi, H.G. (2017). Discrepancy between self-assessed hearing status and measured audiometric evaluation. PLoS One *12*, e0182718.

279. Nondahl, D.M., Cruickshanks, K.J., Wiley, T.L., Tweed, T.S., Klein, R., and Klein, B.E. Accuracy of self-reported hearing loss. Audiology *37*, 295–301.

280. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet. *102*, 717–730.

281. van Rheenen, W., Peyrot, W.J., Schork, A.J., Lee, S.H., and Wray, N.R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. Nat. Rev. Genet. *20*, 567–581.

282. Kraft, P., Zeggini, E., and Ioannidis, J.P.A. (2009). Replication in genome-wide association studies. Stat. Sci. *24*, 561–573.

283. Churchhouse, C., and Neale, B. (2017). Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank.

284. Aschard, H., Vilhjálmsson, B.J., Joshi, A.D., Price, A.L., and Kraft, P. (2015). Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am. J. Hum. Genet. *96*, 329–339.

285. Gates, G.A., Couropmitrcc, N.N., and Alycrs, R.H. (1999). Genetic associations in age-related hearing thresholds. Arch. Otolaryngol. - Head Neck Surg. *125*, 654–659.

286. Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. Hum. Genet. *131*, 1655–1664.

287. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

288. Schönberger, J., Wang, L., Shin, J.T., Sang, D.K., Depreux, F.F.S., Zhu, H., Zon, L., Pizard, A., Kim, J.B., MacRae, C.A., et al. (2005). Mutation in the transcriptional coactivator EYA4 causes dilated cardiomyopathy and sensorineural hearing loss. Nat. Genet. *37*, 418–422.

289. Pfister, M., Tóth, T., Thiele, H., Haack, B., Blin, N., Zenner, H.-P., Sziklai, I., Nürnberg, P., and Kupka, S. (2002). A 4bp-Insertion in the eya-Homologous Region (eyaHR) of EYA4 Causes Hearing Impairment in a Hungarian Family Linked to DFNA10. Mol. Med. *8*, 607–611.

290. Bondurand, N., Dastot-Le Moal, F., Stanchina, L., Collot, N., Baral, V., Marlin, S., Attie-Bitach, T., Giurgea, I., Skopinski, L., Reardon, W., et al. (2007). Deletions at the SOX10 Gene Locus Cause Waardenburg Syndrome Types 2 and 4. Am. J. Hum. Genet. *81*, 1169–1185.

291. Bolz, H., Von Brederlow, B., Ramírez, A., Bryda, E.C., Kutsche, K., Nothwang, H.G., Seeliger, M., Cabrera, M.D.C.S., Vila, M.C., Molina, O.P., et al. (2001). Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. Nat. Genet. *27*, 108–112.

292. Schrauwen, I., Chakchouk, I., Liaqat, K., Jan, A., Nasir, A., Hussain, S., Nickerson, D.A., Bamshad, M.J., Ullah, A., Ahmad, W., et al. (2018). A variant in LMX1A causes autosomal recessive severe-to-profound hearing impairment. Hum. Genet. *137*, 471–478.

293. Sheets, L., Trapani, J.G., Mo, W., Obholzer, N., and Nicolson, T. (2011). Ribeye is required for presynaptic Ca V 1.3a channel localization and afferent innervation of sensory hair cells . Development *138*, 1309–1319.

294. Uthaiah, R.C., and Hudspeth, A.J. (2010). Molecular Anatomy of the Hair Cell's Ribbon Synapse. J. Neurosci. *30*, 12387–12399.

295. Planchart, A. (2013). Analysis of an intronic promoter within Synj2. Biochem. Biophys. Res. Commun. *440*, 640–645.

296. Liu, Y., Qi, J., Chen, X., Tang, M., Chu, C., Zhu, W., Li, H., Tian, C., Yang, G., Zhong, C., et al. (2019). Critical role of spectrin in hearing development and deafness. Sci. Adv. *5*, eaav7803.

297. Ohlemiller, K.K., Hughes, R.M., Mosinger-Ogilvie, J., Speck, J.D., Grosof, D.H., and Silverman, M.S. (1995). Cochlear and retinal degeneration in the Tubby mouse. Neuroreport *6*, 845–849.

298. Ohlemiller, K.K., Hughes, R.M., Lett, J.M., Ogilvie, J.M., Speck, J.D., Wright, J.S., and Faddis, B.T. (1997). Progression of cochlear and retinal degeneration in the tubby (rd5) mouse. Audiol. Neuro-Otology *2*, 175–185.

299. Helfert, R.H., Sommer, T.J., Meeks, J., Hofstetter, P., and Hughes, L.F. (1999). Age-related synaptic changes in the central nucleus of the inferior colliculus of Fischer-344 rats. J. Comp. Neurol. *406*, 285–298.

300. Yu, Y.F., Zhai, F., Dai, C.F., and Hu, J.J. (2011). The relationship between age-related hearing loss and synaptic changes in the hippocampus of C57BL/6J mice. Exp. Gerontol. *46*, 716–722.

301. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

302. Gilles, A., Camp, G., Van de Heyning, P., and Fransen, E. (2017). A pilot genome-wide association study identifies potential metabolic pathways involved in tinnitus. Front. Neurosci. *11*, 71.

303. Stranden, M., Solvin, H., Fors, E.A., Getz, L., and Helvik, A.S. (2016). Are persons with fibromyalgia or other musculoskeletal pain more likely to report hearing loss? A HUNT study. BMC Musculoskelet. Disord. *17*, 477.

304. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: Challenges and strategies. Nat. Rev. Genet. *14*, 483–495.

305. Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. *18*, 117–127.

306. Coons, A.H., Creech, H.J., and Jones, R.N. (2013). Immunological Properties of an Antibody Containing a Fluorescent Group. Exp. Biol. Med. *47*, 200–202.

307. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885.

308. Tadjuidje, E., and Hegde, R.S. (2013). The Eyes Absent proteins in development and disease. Cell. Mol. Life Sci. *70*, 1897–1913.

309. Wayne, S. (2001). Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus. Hum. Mol. Genet. *10*, 195–200.

310. Liu, F., Hu, J., Xia, W., Hao, L., Ma, J., Ma, D., and Ma, Z. (2015). Exome sequencing identifies a mutation in EYA4 as a novel cause of autosomal dominant non-syndromic hearing loss. PLoS One *10*, e0126602.

311. Matsuzaki, S., Hosoya, M., Okano, H., Fujioka, M., and Ogawa, K. (2018). Expression pattern of EYA4 in the common marmoset (Callithrix jacchus) cochlea. Neurosci. Lett. *662*, 185–188.

312. Legendre, K., Safieddine, S., Kussel-Andermann, P., Petit, C., and El-Amraoui, A. (2008). II- V

spectrin bridges the plasma membrane and cortical lattice in the lateral wall of the auditory outer hair cells. J. Cell Sci. *121*, 3347–3356.

313. Cortese, M., Papal, S., Pisciottano, F., Elgoyhen, A.B., Hardelin, J.-P., Petit, C., Franchini, L.F., and El-Amraoui, A. (2017). Spectrin βV adaptive mutations and changes in subcellular location correlate with emergence of hair cell electromotility in mammalians. Proc. Natl. Acad. Sci. *114*, 2054–2059.

314. Kitajiri, S.I., Sakamoto, T., Belyantseva, I.A., Goodyear, R.J., Stepanyan, R., Fujiwara, I., Bird, J.E., Riazuddin, S., Riazuddin, S., Ahmed, Z.M., et al. (2010). Actin-bundling protein TRIOBP forms resilient rootlets of hair cell stereocilia essential for hearing. Cell *141*, 786–798.

315. Boggon, T.J., Shan, W.S., Santagata, S., Myers, S.C., and Shapiro, L. (1999). Implication of tubby proteins as transcription factors by structure-based functional analysis. Science *286*, 2119–2125.

316. Ohlemiller, K.K., Hughes, R.M., Lett, J.M., Ogilvie, J.M., Speck, J.D., Wright, J.S., and Faddis, B.T. (1997). Progression of cochlear and retinal degeneration in the tubby (rd5) mouse. Audiol. Neuro-Otology *2*, 175–185.

317. Yousaf, R., Friedman, T.B., and Riazuddin, S. (2016). Genetic modifiers of hearing loss. In Monographs in Human Genetics, (Karger Publishers), pp. 73–83.

318. Manji, S.S.M., Williams, L.H., Miller, K.A., Ooms, L.M., Bahlo, M., Mitchell, C.A., and Dahl, H.H.M. (2011). A mutation in synaptojanin 2 causes progressive hearing loss in the ENU-mutagenised mouse strain mozart. PLoS One *6*, e17607.

319. Gasparoni, G., Bultmann, S., Lutsik, P., Kraus, T.F.J., Sordon, S., Vlcek, J., Dietinger, V., Steinmaurer, M., Haider, M., Mulholland, C.B., et al. (2018). DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics and Chromatin *11*, 41.

320. Lopez, L.M., Harris, S.E., Luciano, M., Liewald, D., Davies, G., Gow, A.J., Tenesa, A., Payton, A., Ke, X., Whalley, L.J., et al. (2012). Evolutionary conserved longevity genes and human cognitive abilities in elderly cohorts. Eur. J. Hum. Genet. *20*, 341–347.

321. Talebi, F., Mardasi, F.G., Asl, J.M., and Sayahi, M. (2017). Next-generation sequencing identifies three novel missense variants in ILDR1 and MYO6 genes in an Iranian family with hearing loss with review of the literature. Int. J. Pediatr. Otorhinolaryngol. *103*, 103–108.

322. Wang, X., Wang, L., Peng, H., Yang, T., and Wu, H. (2018). A Novel p.G141R Mutation in ILDR1 Leads to Recessive Nonsyndromic Deafness DFNB42 in Two Chinese Han Families. Neural Plast. *2018*, 1–6.

323. Diaz-Horta, O., Duman, D., Foster, J., Sirmaci, A., Gonzalez, M., Mahdieh, N., Fotouhi, N., Bonyadi, M., Cengiz, F.B., Menendez, I., et al. (2012). Whole-Exome Sequencing Efficiently Detects Rare Mutations in Autosomal Recessive Nonsyndromic Hearing Loss. PLoS One *7*, e50628.

324. Borck, G., Rehman, A.U., Lee, K., Pogoda, H.M., Kakar, N., Von Ameln, S., Grillet, N., Hildebrand, M.S., Ahmed, Z.M., Nürnberg, G., et al. (2011). Loss-of-function mutations of ILDR1 cause autosomal-recessive hearing impairment DFNB42. Am. J. Hum. Genet. *88*, 127–137.

325. Sang, Q., Li, W., Xu, Y., Qu, R., Xu, Z., Feng, R., Jin, L., He, L., Li, H., and Wang, L. (2015). ILDR1 deficiency causes degeneration of cochlear outer hair cells and disrupts the structure of the organ of Corti: a mouse model for human DFNB42. Biol. Open *4*, 411–418.

326. Kitajiri, S.I., and Katsuno, T. (2016). Tricellular Tight Junctions in the Inner Ear. Biomed Res. Int. *2016*, 1–5.

327. Kim, N.K.D., Higashi, T., Lee, K.Y., Kim, A.R., Kitajiri, S., Kim, M.Y., Chang, M.Y., Kim, V., Oh, S.H., Kim, D., et al. (2015). Downsloping high-frequency hearing loss due to inner ear tricellular tight junction disruption by a novel ILDR1 mutation in the Ig-like domain. PLoS One *10*, e0116931.

328. Steffes, G., Lorente-Cánovas, B., Pearson, S., Brooker, R.H., Spiden, S., Kiernan, A.E., Guénet, J.L., and Steel, K.P. (2012). Mutanlallemand (mtl) and Belly Spot and Deafness (bsd) Are Two New Mutations of Lmx1a Causing Severe Cochlear and Vestibular Defects. PLoS One *7*, e51065.

329. Jaiganesh, A., De-la-Torre, P., Patel, A.A., Termine, D.J., Velez-Cortes, F., Chen, C., and Sotomayor, M. (2018). Zooming in on Cadherin-23: Structural Diversity and Potential Mechanisms of Inherited Deafness. Structure *26*, 1210-1225.e4.

330. Bouzid, A., Smeti, I., Chakroun, A., Loukil, S., Gibriel, A.A., Grati, M., Ghorbel, A., and Masmoudi, S. (2018). CDH23 methylation status and presbycusis risk in elderly women. Front. Aging Neurosci. *10*, 241.

331. Liang, Y., Wang, K., Peng, Q., Zhu, P., Wu, C., Rao, C., Chang, J., Li, S., and Lu, X. (2018). A novel variant in the CDH23 gene is associated with non-syndromic hearing loss in a Chinese family. Int. J. Pediatr. Otorhinolaryngol. *104*, 108–112.

332. Ganapathy, A., Pandey, N., Srisailapathy, C.R.S., Jalvi, R., Malhotra, V., Venkatappa, M., Chatterjee, A., Sharma, M., Santhanam, R., Chadha, S., et al. (2014). Non-syndromic hearing impairment in India: High allelic heterogeneity among mutations in TMPRSS3, TMC1, USHIC, CDH23 and TMIE. PLoS One *9*, e84773.

333. Xu, T., Zhu, W., and Wang, P. (2019). The p.P240L variant of CDH23 and the risk of nonsyndromic hearing loss: a meta-analysis. Eur. Arch. Oto-Rhino-Laryngology *276*, 11–16.

334. Magupalli, V.G., Schwarz, K., Alpadi, K., Natarajan, S., Seigel, G.M., and Schmitz, F. (2008). Multiple RIBEYE-RIBEYE Interactions Create a Dynamic Scaffold for the Formation of Synaptic Ribbons. J. Neurosci. *28*, 7954–7967.

335. Jean, P., de la Morena, D.L., Michanski, S., Tobón, L.M.J., Chakrabarti, R., Picher, M.M., Neef, J., Jung, S.Y., Gültas, M., Maxeiner, S., et al. (2018). The synaptic ribbon is critical for sound encoding at high

rates and with temporal precision. Elife *7*, e29275.

336. Depreux, F.F.S., Darrow, K., Conner, D.A., Eavey, R.D., Liberman, M.C., Seidman, C.E., and Seidman, J.G. (2008). Eya4-deficient mice are a model for heritable otitis media. J. Clin. Invest. *118*, 651–658.

337. Ikeda, A., Zheng, Q.Y., Zuberi, A.R., Johnson, K.R., Naggert, J.K., and Nishina, P.M. (2002). Microtubule-associated protein 1A is a modifier of tubby hearing (moth1). Nat. Genet. *30*, 401–405.

338. Sang, Q., Li, W., Xu, Y., Qu, R., Xu, Z., Feng, R., Jin, L., He, L., Li, H., and Wang, L. (2015). ILDR1 deficiency causes degeneration of cochlear outer hair cells and disrupts the structure of the organ of Corti: a mouse model for human DFNB42. Biol. Open *4*, 411–418.

339. Huang, Y., Hill, J., Yatteau, A., Wong, L., Jiang, T., Petrovic, J., Gan, L., Dong, L., and Wu, D.K. (2018). Reciprocal Negative Regulation Between Lmx1a and Lmo4 Is Required for Inner Ear Formation. J. Neurosci. *38*, 5429–5440.

340. Kazmierczak, P., Sakaguchi, H., Tokita, J., Wilson-Kubalek, E.M., Milligan, R.A., Müller, U., and Kachar, B. (2007). Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. Nature *449*, 87–91.

341. Sheets, L., Trapani, J.G., Mo, W., Obholzer, N., and Nicolson, T. (2011). Ribeye is required for presynaptic Ca V 1.3a channel localization and afferent innervation of sensory hair cells . Development *138*, 1309–1319.

342. Koran, M.E.I., Hohman, T.J., Meda, S.A., and Thornton-Wells, T.A. (2014). Genetic interactions within inositol-related pathways are associated with longitudinal changes in ventricle size. J. Alzheimer's Dis. *38*, 145–154.

343. Tamási, V., Petschner, P., Adori, C., Kirilly, E., Ando, R.D., Tothfalusi, L., Juhasz, G., and Bagdy, G. (2014). Transcriptional evidence for the role of chronic venlafaxine treatment in neurotrophic signaling and neuroplasticity including also glutatmatergic- and insulin-mediated neuronal processes. PLoS One *9*, e113662.

344. Luciano, M., Lopez, L.M., De Moor, M.H.M., Harris, S.E., Davies, G., Nutile, T., Krueger, R.F., Esko, T., Schlessinger, D., Toshiko, T., et al. (2012). Longevity candidate genes and their association with personality traits in the elderly. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. *159 B*, 192–200.

345. Scheffer, D.I., Shen, J., Corey, D.P., and Chen, Z.Y. (2015). Gene expression by mouse inner ear hair cells during development. J. Neurosci. *35*, 6366–6380.

346. Feng, H., Xia, X., Li, C., Song, Y., Qin, C., Zhang, Y., and Lan, X. (2015). TYR as a multifunctional reporter gene regulated by the Tet-on system for multimodality imaging: An in vitro study. Sci. Rep. *5*, 15502.

347. Uehara, S., Izumi, Y., Kubo, Y., Wang, C.C., Mineta, K., Ikeo, K., Gojobori, T., Tachibana, M., Kikuchi, T., Kobayashi, T., et al. (2009). Specific expression of Gsta4 in mouse cochlear melanocytes: A novel

role for hearing and melanocyte differentiation. Pigment Cell Melanoma Res. *22*, 111–119.

348. Murphy, P.A., Begum, S., and Hynes, R.O. (2015). Tumor angiogenesis in the absence of fibronectin or its cognate integrin receptors. PLoS One *10*, e0120872.

349. Dunbar, L., Patni, P., Aguilar, C., Mburu, P. Corns, L., Wells, H.R.R., Delmaghani, S., Parker, A. Johnson, S., Chessum, L., et al. (2019). Clarin-2 is Required for Maintenance of Mechanotransducing Stereocilia and Essential for Hearing. In ARO Midwinter Meeting Abstract, p. https://cdn.ymaws.com/www.aro.org/resource/resmgr/.

350. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

351. Pircs, K., Petri, R., Madsen, S., Brattås, P.L., Vuono, R., Ottosson, D.R., St-Amour, I., Hersbach, B.A., Matusiak-Brückner, M., Lundh, S.H., et al. (2018). Huntingtin Aggregation Impairs Autophagy, Leading to Argonaute-2 Accumulation and Global MicroRNA Dysregulation. Cell Rep. *24*, 1397–1406.

352. Gao, Y., Liu, X.F., Lu, X.C., Ma, C., Cao, J., and Fan, L. (2012). Protective effects of atorvastatin against oxidized LDL-induced downregulation of KLF expression in EA.hy926 cells. Int. J. Mol. Med. *30*, 330–336.

353. Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cristea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. Mol. Syst. Biol. *9*, 672.

354. Vlajkovic, S.M., Guo, C.X., Telang, R., Wong, A.C.Y., Paramananthasivam, V., Boison, D., Housley, G.D., and Thorne, P.R. (2011). Adenosine kinase inhibition in the cochlea delays the onset of age-related hearing loss. Exp. Gerontol. *46*, 905–914.

355. Ren, A., Zhang, W., Yarlagadda, S., Sinha, C., Arora, K., Moon, C.S., and Naren, A.P. (2013). MAST205 competes with cystic fibrosis transmembrane conductance regulator (CFTR)-associated ligand for binding to CFTR to regulate CFTR-mediated fluid transport. J. Biol. Chem. *288*, 12325–12334.

356. Nielsen, J. V., Nielsen, F.H., Ismail, R., Noraberg, J., and Jensen, N.A. (2007). Hippocampus-like corticoneurogenesis induced by two isoforms of the BTB-zinc finger gene Zbtb20 in mice. Development *134*, 1133–1140.

357. Tan, J., Yang, L., Liu, C., and Yan, Z. (2017). MicroRNA-26a targets MAPK6 to inhibit smooth muscle cell proliferation and vein graft neointimal hyperplasia. Sci. Rep. *7*, 46602.

358. Sipe, C.W., Liu, L., Lee, J., Grimsley-Myers, C., and Lu, X. (2013). Lis1 mediates planar polarity of auditory hair cells through regulation of microtubule organization. Dev. *140*, 1785–1795.

359. Lucas, M.E. (2006). The bHLH transcription factor hand2 is essential for noradrenergic differentiation of sympathetic neurons. Development *133*, 4015–4024.

360. Gnedeva, K., and Hudspeth, A.J. (2015). SoxC transcription factors are essential for the development of the inner ear. Proc. Natl. Acad. Sci. U. S. A. *112*, 14066–14071.

361. Mathioudakis, N., Sundaresh, R., Larsen, A., Ruff, W., Schiller, J., Guerrero-Cázares, H., Burger, P., Salvatori, R., and Quiñones-Hinojosa, A. (2015). Expression of the pituitary stem/progenitor marker GFRα2 in human pituitary adenomas and normal pituitary. Pituitary *18*, 31–41.

362. Liu, W., and Rask-Andersen, H. (2014). Immunohistological analysis of neurturin and its receptors in human cochlea. Auris Nasus Larynx *41*, 172–178.

363. Liu, W., Li, H., Hong, S.H., Piszczek, G.P., Chen, W., and Rodgers, G.P. (2016). Olfactomedin 4 deletion induces colon adenocarcinoma in ApcMin/+ mice. Oncogene *35*, 5237–5247.

364. Kutsche, K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H.G., Orth, U., Boavida, M.G., David, D., Chelly, J., Fryns, J.P., et al. (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. Nat. Genet. *26*, 247–250.

365. Zhu, C., Cheng, C., Wang, Y., Muhammad, W., Liu, S., Zhu, W., Shao, B., Zhang, Z., Yan, X., He, Q., et al. (2018). Loss of ARHGEF6 Causes Hair Cell Stereocilia Deficits and Hearing Loss in Mice. Front. Mol. Neurosci. *11*, 362.

366. Ishiyama, A., Mowry, S.E., Lopez, I.A., and Ishiyama, G. (2009). Immunohistochemical distribution of basement membrane proteins in the human inner ear from older subjects. Hear. Res. *254*, 1–14.

367. Fields, R.R., Zhou, G., Huang, D., Davis, J.R., Möller, C., Jacobson, S.G., Kimberling, W.J., and Sumegi, J. (2002). Usher Syndrome Type III: Revised Genomic Structure of the USH3 Gene and Identification of Novel Mutations. Am. J. Hum. Genet. *71*, 607–617.

368. Adato, A., Vreugde, S., Joensuu, T., Avidan, N., Hamalainen, R., Belenkiy, O., Olender, T., Bonne-Tamir, B., Ben-Asher, E., Espinos, C., et al. (2002). USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. Eur. J. Hum. Genet. *10*, 339–350.

369. Dunbar, L.A., Patni, P., Aguilar, C., Mburu, P., Corns, L., Wells, H.R., Delmaghani, S., Parker, A., Johnson, S., Williams, D., et al. (2019). Clarin-2 is essential for hearing by maintaining stereocilia integrity and function. EMBO Mol. Med. *11*, e10288.

370. Tarchini, B., Longo-Guess, C., Tian, C., Tadenev, A.L.D., Devanney, N., and Johnson, K.R. (2018). A spontaneous mouse deletion in Mctp1 uncovers a long-range cis-regulatory region crucial for NR2F1 function during inner ear development. Dev. Biol. *443*, 153–164.

371. Hickox, A.E., Wong, A.C.Y., Pak, K., Strojny, C., Ramirez, M., Yates, J.R., Ryan, A.F., and Savas, J.N. (2017). Global analysis of protein expression of inner ear hair cells. J. Neurosci. *37*, 1320–1339.

372. Kohfeldt, E., Sasaki, T., Göhring, W., and Timpl, R. (1998). Nidogen-2: A new basement membrane

protein with diverse binding properties.

373. Schymeinsky, J., Nedbal, S., Miosge, N., Poschl, E., Rao, C., Beier, D.R., Skarnes, W.C., Timpl, R., and Bader, B.L. (2002). Gene Structure and Functional Analysis of the Mouse Nidogen-2 Gene: Nidogen-2 Is Not Essential for Basement Membrane Formation in Mice. Mol. Cell. Biol. *22*, 6820–6830.

374. Seong Hoon Kang, and Kramer, J.M. (2000). Nidogen is nonessential and not required for normal type IV collagen localization in Caenorhabditis elegans. Mol. Biol. Cell *11*, 3911–3923.

375. Bader, B.L., Smyth, N., Nedbal, S., Miosge, N., Baranowsky, A., Mokkapati, S., Murshed, M., and Nischt, R. (2005). Compound Genetic Ablation of Nidogen 1 and 2 Causes Basement Membrane Defects and Perinatal Lethality in Mice. Mol. Cell. Biol. *25*, 6846–6856.

376. Van Horck, F.P.G., Ahmadian, M.R., Haeusler, L.C., Moolenaar, W.H., and Kranenburg, O. (2001). Characterization of p190RhoGEF, A RhoA-specific Guanine Nucleotide Exchange Factor That Interacts with Microtubules. J. Biol. Chem. *276*, 4948–4956.

377. Droppelmann, C.A., Wang, J., Campos-Melo, D., Keller, B., Volkening, K., Hegele, R.A., and Strong, M.J. (2013). Detection of a novel frameshift mutation and regions with homozygosis within ARHGEF28 gene in familial amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Front. Degener. *14*, 444–451.

378. Farhan, S.M.K., Gendron, T.F., Petrucelli, L., Hegele, R.A., and Strong, M.J. (2017). ARHGEF28 P.LYS280METFS40TER in an amyotrophic lateral sclerosis family with a C9ORF72 expansion. Neurol. Genet. *3*, e190.

379. Ma, Y., Tang, L., Chen, L., Zhang, B., Deng, P., Wang, J., Yang, Y., Liu, R., Yang, Y., Ye, S., et al. (2014). ARHGEF28 gene exon 6/intron 6 junction mutations in Chinese amyotrophic lateral sclerosis cohort. Amyotroph. Lateral Scler. Front. Degener. *15*, 309–311.

380. Song, Y., Lin, F., Ye, C. hui, Huang, H., Li, X., Yao, X., Xu, Y., and Wang, C. (2019). Rare, low-frequency and common coding variants of ARHGEF28 gene and their association with sporadic amyotrophic lateral sclerosis. Neurobiol. Aging.

381. Previtali, S.C., Zhao, E., Lazarevic, D., Pipitone, G.B., Fabrizi, G.M., Manganelli, F., Mazzeo, A., Pareyson, D., Schenone, A., Taroni, F., et al. (2019). Expanding the spectrum of genes responsible for hereditary motor neuropathies. J. Neurol. Neurosurg. Psychiatry jnnp-2019-320717.

382. Abiko, H., Fujiwara, S., Ohashi, K., Hiatari, R., Mashiko, T., Sakamoto, N., Sato, M., and Mizuno, K. (2015). Rho guanine nucleotide exchange factors involved in cyclic-stretch-induced reorientation of vascular endothelial cells. J. Cell Sci. *128*, 1683–1695.

383. Zaritsky, A., Tseng, Y.Y., Rabadán, M.A., Krishna, S., Overholtzer, M., Danuser, G., and Hall, A. (2017). Diverse roles of guanine nucleotide exchange factors in regulating collective cell migration. J. Cell Biol. *216*, 1543–1556.

384. Miller, N.L.G., Kleinschmidt, E.G., and Schlaepfer, D.D. (2014). RhoGEFs in Cell Motility: Novel Links Between Rgnef and Focal Adhesion Kinase. Curr. Mol. Med. *14*, 221–234.

385. Pisamai, S., Roytrakul, S., Phaonakrop, N., Jaresitthikunchai, J., and Suriyaphol, G. (2018). Proteomic analysis of canine oral tumor tissues using MALDI-TOF mass spectrometry and in-gel digestion coupled with mass spectrometry (GeLC MS/MS) approaches. PLoS One *13*, e0200619.

386. Kleinschmidt, E.G., Miller, N.L.G., Ozmadenci, D., Tancioni, I., Osterman, C.D., Barrie, A.M., Taylor, K.N., Ye, A., Jiang, S., Connolly, D.C., et al. (2019). Rgnef promotes ovarian tumor progression and confers protection from oxidative stress. Oncogene.

387. Masià-Balagué, M., Izquierdo, I., Garrido, G., Cordomí, A., Pérez-Benito, L., Miller, N.L.G., Schlaepfer, D.D., Gigoux, V., and Aragay, A.M. (2015). Gastrin-stimulated Gα<inf>13</inf> activation of Rgnef protein (ArhGEF28) in DLD-1 colon carcinoma cells. J. Biol. Chem. *290*, 15197–15209.

388. Zallocchi, M., Meehan, D.T., Delimont, D., Askew, C., Garige, S., Gratton, M.A., Rothermund-Franklin, C.A., and Cosgrove, D. (2009). Localization and expression of clarin-1, the Clrn1 gene product, in auditory hair cells and photoreceptors. Hear. Res. *255*, 109–120.

389. Geng, R., Geller, S.F., Hayashi, T., Ray, C.A., Reh, T.A., Bermingham-McDonogh, O., Jones, S.M., Wright, C.G., Melki, S., Imanishi, Y., et al. (2009). Usher syndrome IIIA gene clarin-1 is essential for hair cell function and associated neural activation. Hum. Mol. Genet. *18*, 2748–2760.

390. Geng, R., Melki, S., Chen, D.H.-C., Tian, G., Furness, D.N., Oshima-Takago, T., Neef, J., Moser, T., Askew, C., Horwitz, G., et al. (2012). The Mechanosensory Structure of the Hair Cell Requires Clarin-1, a Protein Encoded by Usher Syndrome III Causative Gene. J. Neurosci. *32*, 9485–9498.

391. Lukashkin, A.N., Richardson, G.P., and Russell, I.J. (2010). Multiple roles for the tectorial membrane in the active cochlea. Hear. Res. *266*, 26–35.

392. Salt, A.N., Melichar, I., and Thalmann, R. (1987). Mechanisms of Endocochlear Potential Generation By Stria Vascularis. Laryngoscope *97*, 984–991.

393. Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, Trans-Omics for Precision Medicine Consortium, Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., et al. (2019). Recovery of trait heritability from whole genome sequence data. BioRxiv 588020.

394. Deary, V., Hagenaars, S.P., Harris, S.E., Hill, W.D., Davies, G., Liewald, D.C.M., McIntosh, A.M., Gale, C.R., and Deary, I.J. (2018). Genetic contributions to self-reported tiredness. Mol. Psychiatry *23*, 609–620.

395. Lane, J.M., Vlasac, I., Anderson, S.G., Kyle, S.D., Dixon, W.G., Bechtold, D.A., Gill, S., Little, M.A., Luik, A., Loudon, A., et al. (2016). Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. Nat. Commun. *7*, 10889.

396. Clarke, T.K., Adams, M.J., Davies, G., Howard, D.M., Hall, L.S., Padmanabhan, S., Murray, A.D.,

Smith, B.H., Campbell, A., Hayward, C., et al. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112117). Mol. Psychiatry *22*, 1376–1384.

397. Davies, N.M., Holmes, M. V., and Davey Smith, G. (2018). Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ *362*, k601.

398. O'Connor, L.J., and Price, A.L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat. Genet. *50*, 1728–1734.

399. Sergeyenko, Y., Lall, K., Charles Liberman, M., and Kujawa, S.G. (2013). Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline. J. Neurosci. *33*, 13686–13694.

400. Koo, S.K., Hill, J.K., Hwang, C.H., Lin, Z.S., Millen, K.J., and Wu, D.K. (2009). Lmx1a maintains proper neurogenic, sensory, and non-sensory domains in the mammalian inner ear. Dev. Biol. *333*, 14–25.

401. Wesdorp, M., de Koning Gans, P.A.M., Schraders, M., Oostrik, J., Huynen, M.A., Venselaar, H., Beynon, A.J., van Gaalen, J., Piai, V., Voermans, N., et al. (2018). Heterozygous missense variants of LMX1A lead to nonsyndromic hearing impairment and vestibular dysfunction. Hum. Genet. *137*, 389–400.

402. Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.., Wendland, J.., Tung, J., Hinds, D.., Perlis, R.., and Winslow, A.. (2017). depression in individuals of European descent. Nat Genet *48*, 1031–1036.

403. Mills, M.C., and Rahal, C. (2019). A scientometric review of genome-wide association studies. Commun. Biol. *2*, 9.

404. Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. Nature *538*, 161–164.

405. Vuckovic, D., Mezzavilla, M., Cocca, M., Morgan, A., Brumat, M., Catamo, E., Concas, M.P., Biino, G., Franzè, A., Ambrosetti, U., et al. (2018). Whole-genome sequencing reveals new insights into age-related hearing loss: cumulative effects, pleiotropy and the role of selection. Eur. J. Hum. Genet. *26*, 1167–1179.

406. Verdi, S., Abbasian, G., Bowyer, R.C.E., Lachance, G., Yarand, D., Christofidou, P., Mangino, M., Menni, C., Bell, J.T., Falchi, M., et al. (2019). TwinsUK: The UK Adult Twin Registry Update. Twin Res. Hum. Genet. 1–7.

407. Banos, D.T., McCartney, D.L., Battram, T., Hemani, G., Walker, R.M., Morris, S.W., Zhang, Q., Porteous, D.J., McRae, A.F., Wray, N.R., et al. (2018). Bayesian reassessment of the epigenetic architecture of complex traits. BioRxiv 450288.

408. Korrapati, S., Taukulis, I., Olszewski, R., Pyle, M., Gu, S., Singh, R., Griffiths, C., Martin, D., Boger, E., Morell, R.J., et al. (2019). Single Cell and Single Nucleus RNA-Seq Reveal Cellular Heterogeneity

and Homeostatic Regulatory Networks in Adult Mouse Stria Vascularis. Front. Mol. Neurosci. *12*, 316.

409. Sugrue, L.P., and Desikan, R.S. (2019). What Are Polygenic Scores and Why Are They Important? JAMA - J. Am. Med. Assoc. *321*, 1820–1821.

410. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.

Appendix

The following locus plots display genetic association results for SNPs located at genome regions flanking all lead SNPs from HDiff and HAid analysis. The following applies to all plots and so is not included in individual legends:

Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r2) to the lead SNP (purple). Where LD information is not available, SNPs are coloured grey. The plots display 200kb +/- regions flanking the position of the lead SNP, except in gene-dense regions, where 100kb +/- is displayed. Two of the lead SNPs are not in the Locuszoom SNP reference database (rs759016271 and 3:182069497_TA_T) and so SNPs in close proximity to these lead SNPs were used to create the locus plots.

Locus plots (Figure a-f) for lead SNPs that reside in exon regions

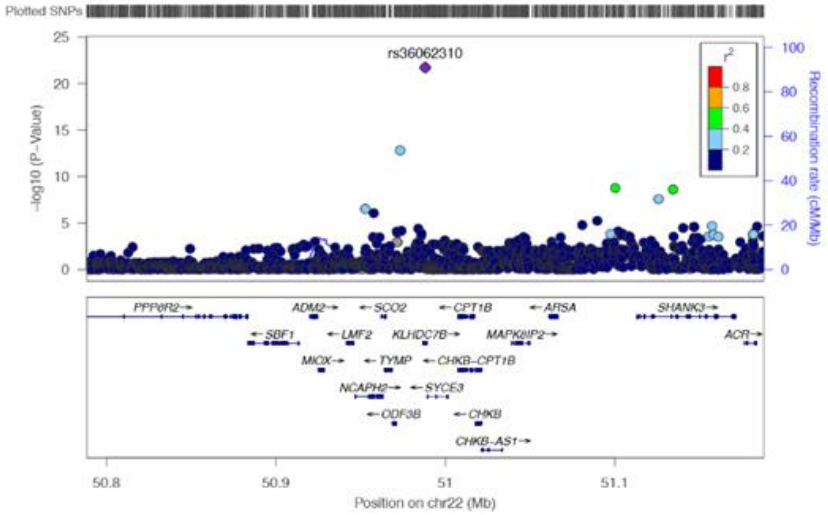

Figure a. Locus plot for genomic region flanking rs36062310

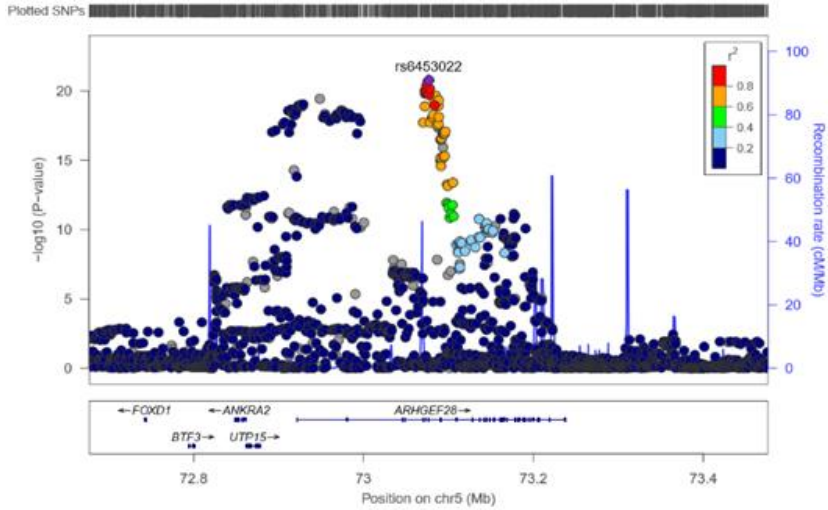

Figure b. Locus plot for genome region flanking SNP rs6453022

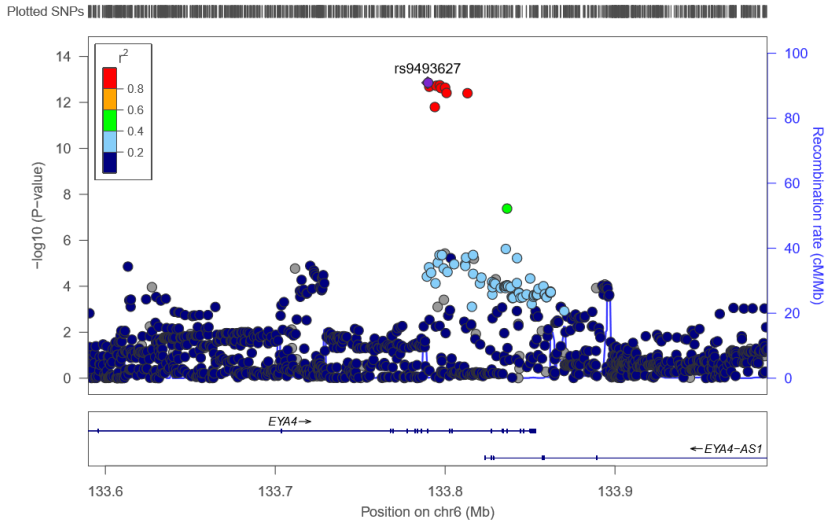

Figure c. Locus plot for genome region flanking SNP rs9493627

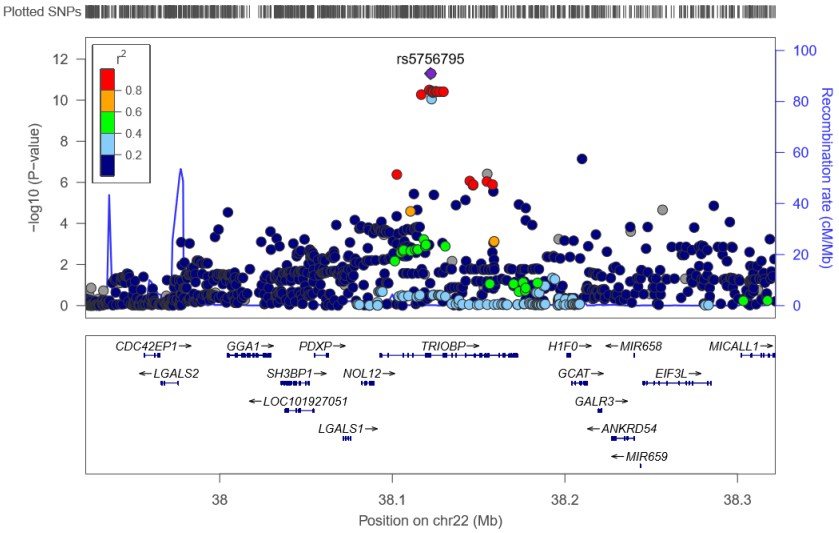

Figure d. Locus plot for region flanking SNP rs5756795

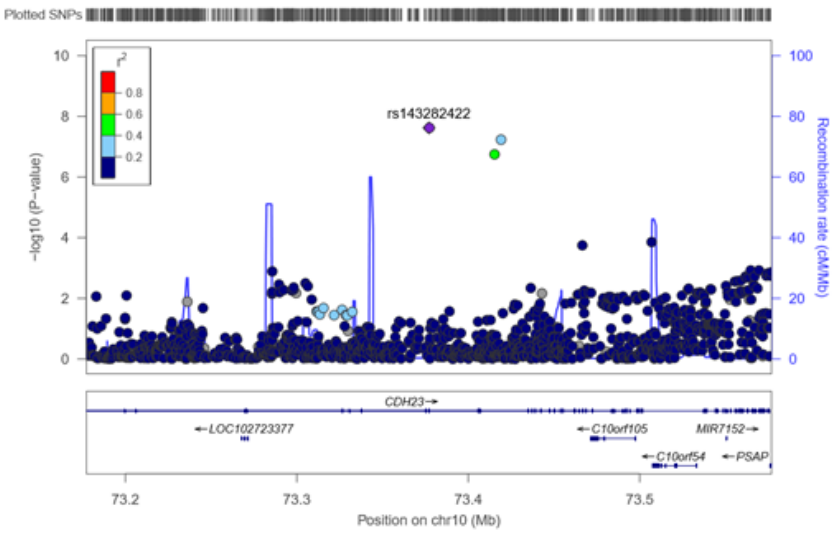

Figure e. Locus plot of genome region flanking rs143282422

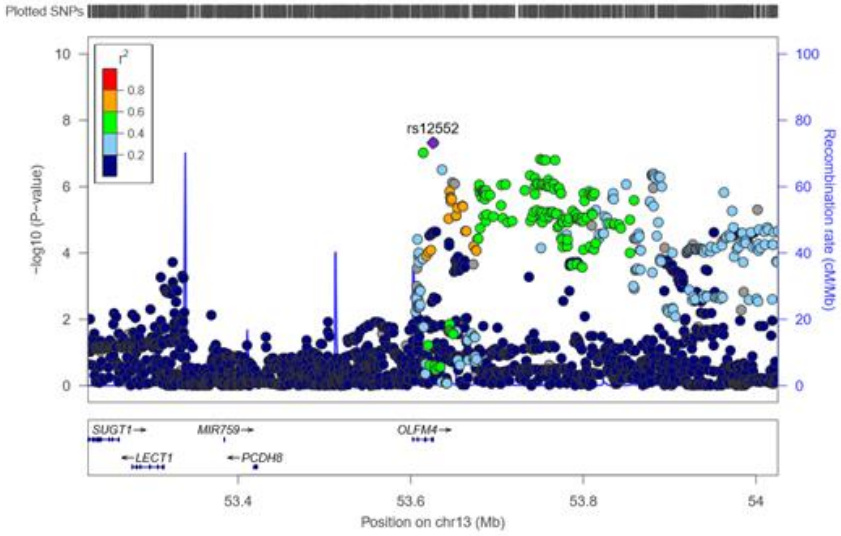

Figure f. Locus plot for genome region flanking SNP rs12552

Locus plots (g-hh) for lead SNPs that reside in intron regions

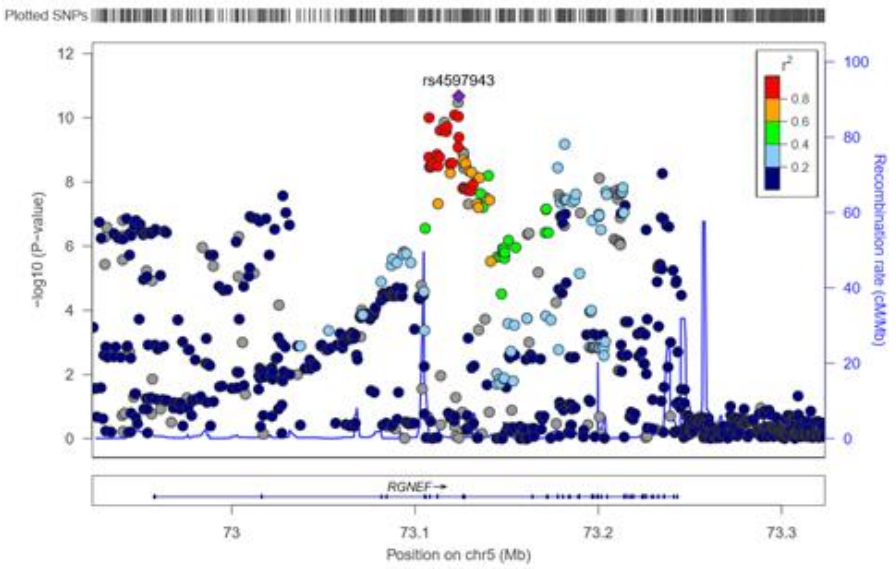

Figure g. Locus plot for genome region flanking SNP rs4597943

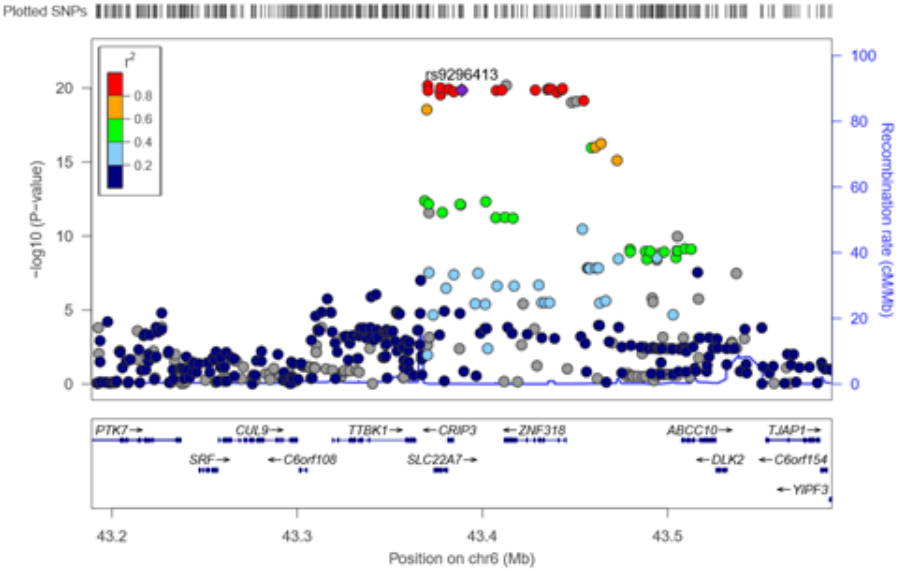

Figure h. Genome region flanking SNP rs9296413

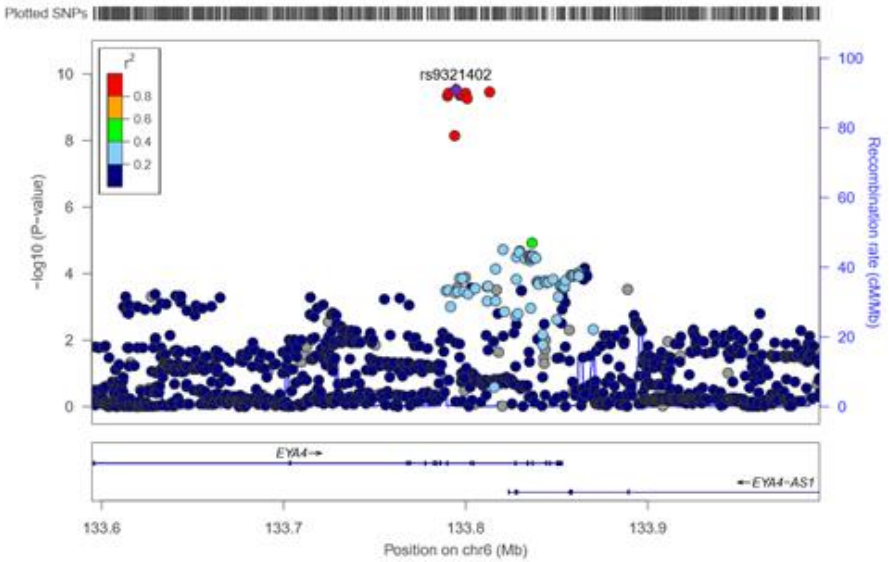

Figure i. Locus plot for genome region flanking rs9321402

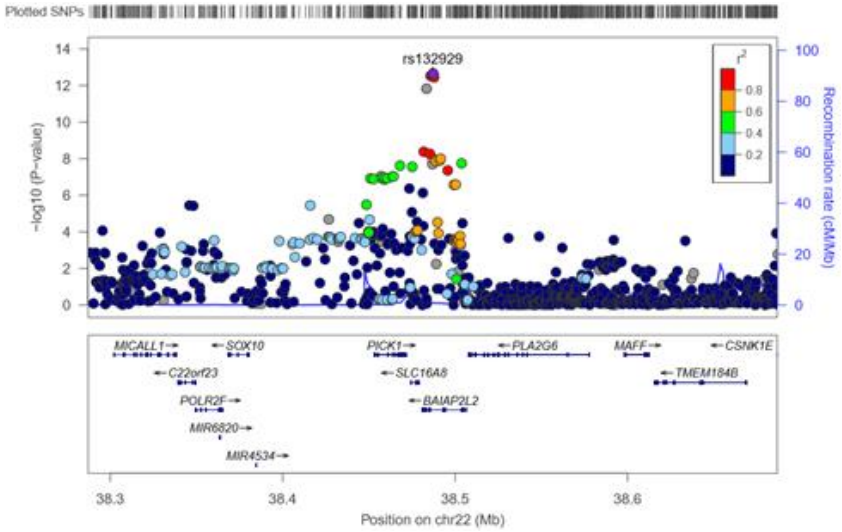

Figure j. Locus plot of genomic region flanking SNP rs132929

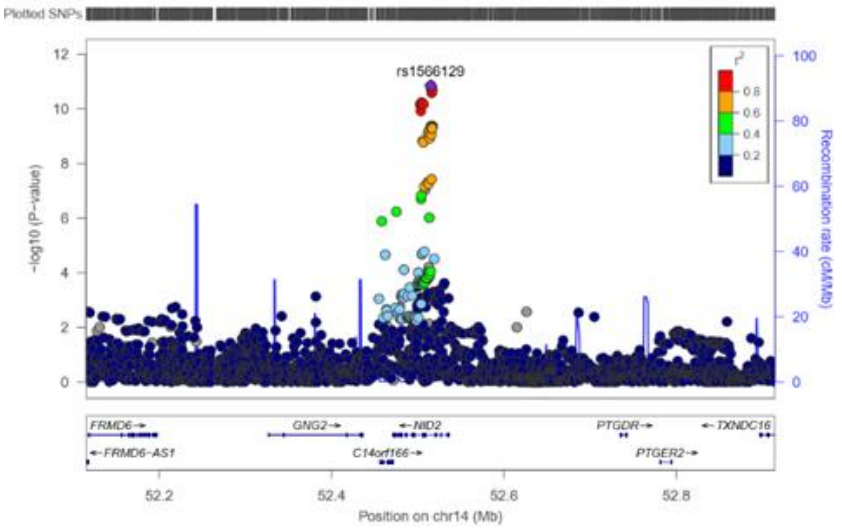

Figure k. Locus plot of genome region flanking SNP rs1566129

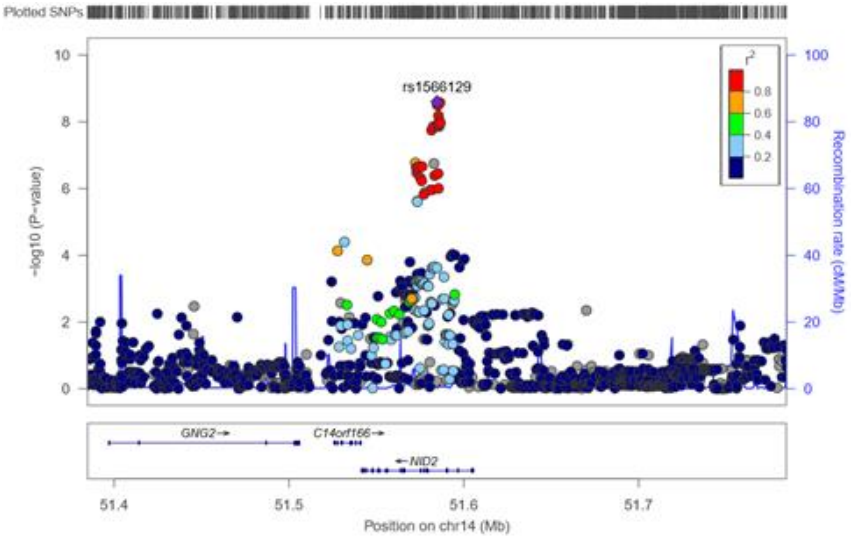

Figure l. Locus plot of genome region flanking SNP rs1566129

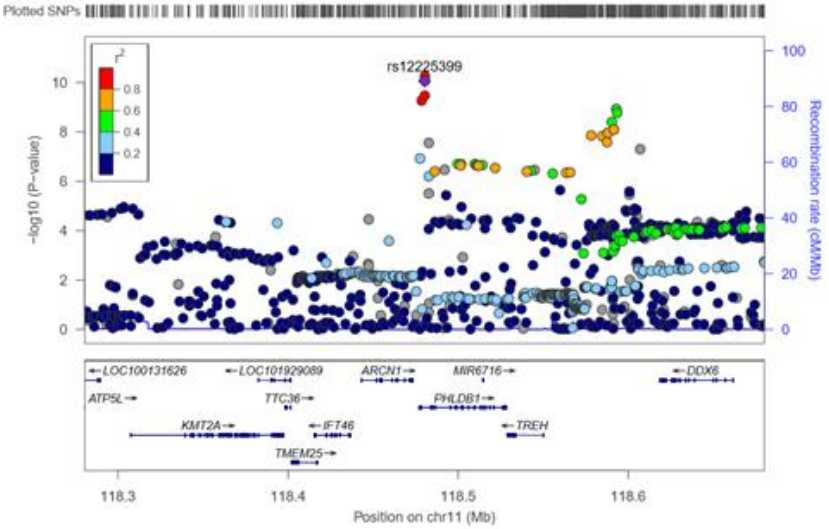

Figure m. Locus plot of genome region flanking rs12225399

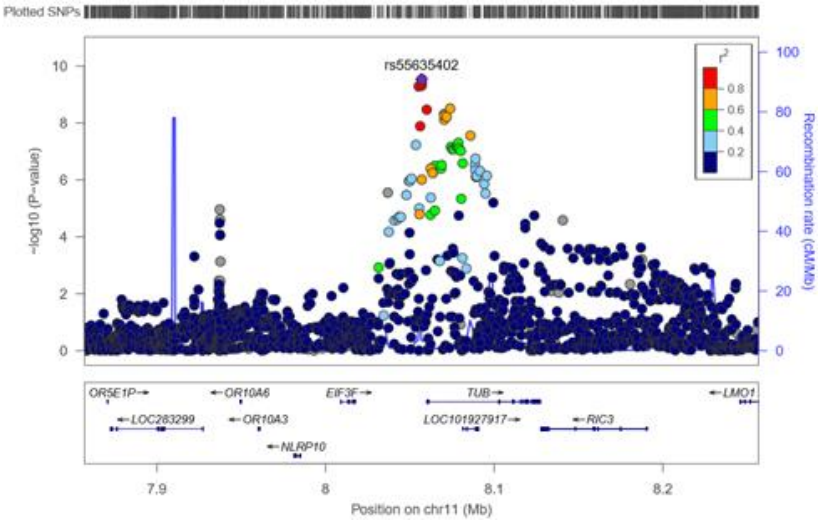

Figure n. Locus plot of genome region flanking SNP rs55635402

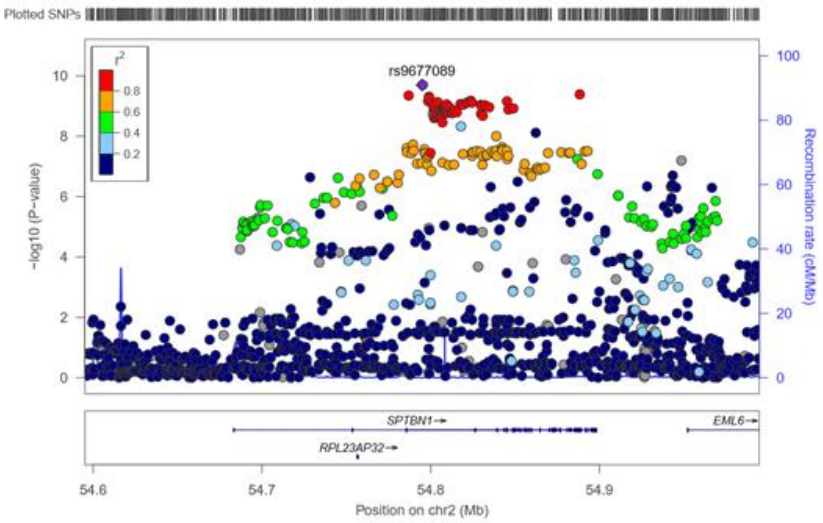

Figure o. Locus plot flanking genome region SNP rs9677089

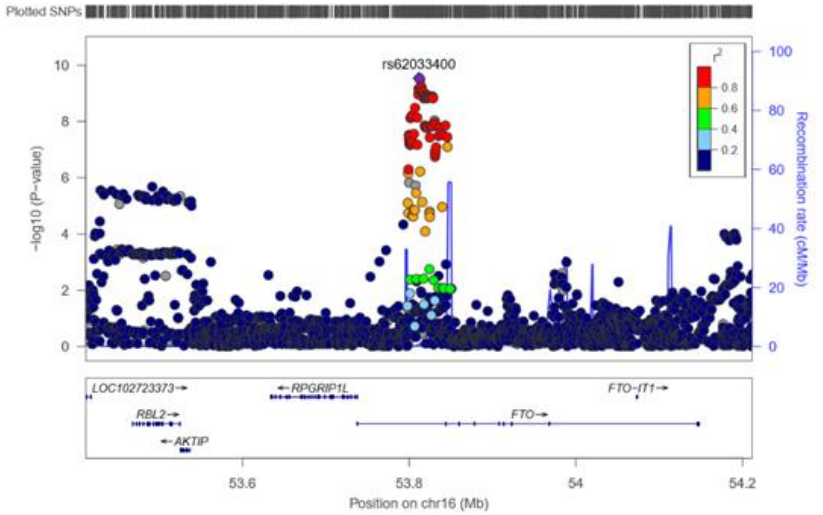

Figure p. Locus plot of genome region flanking SNP rs62033400

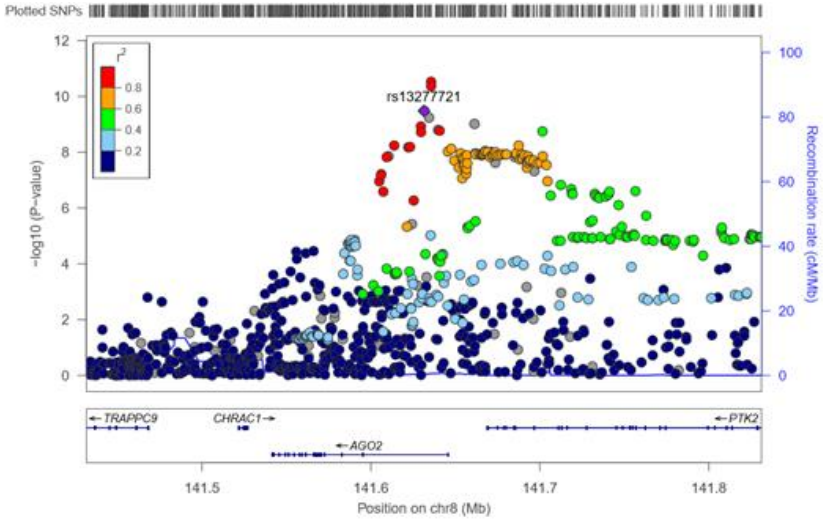

Figure q. Locus plot of genome region flanking SNP rs13277721

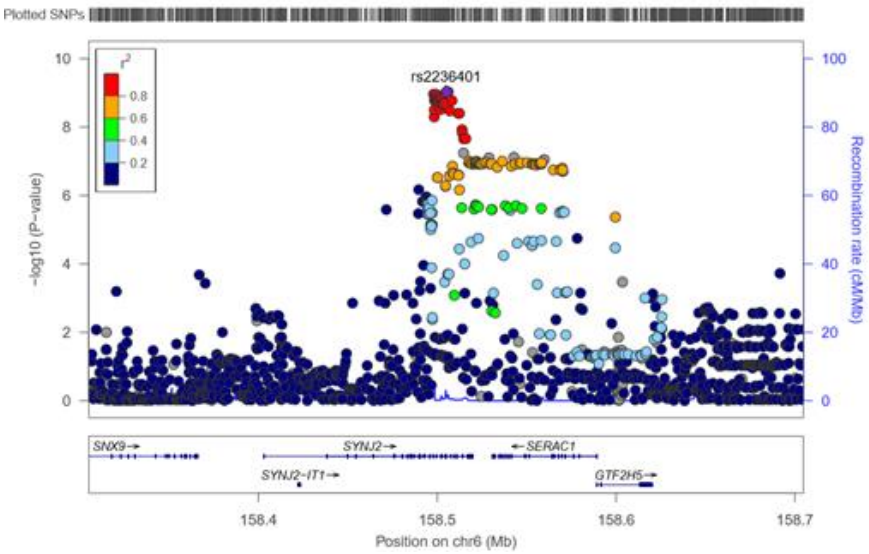

Figure r. Locus plot of genome region flanking SNP rs2236401

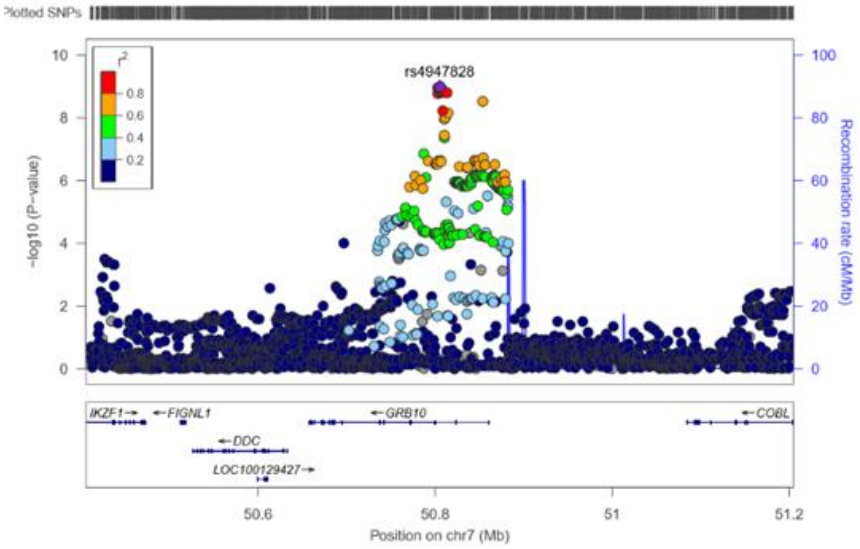

Figure s. Locus plot of genome region flanking SNP rs4947828

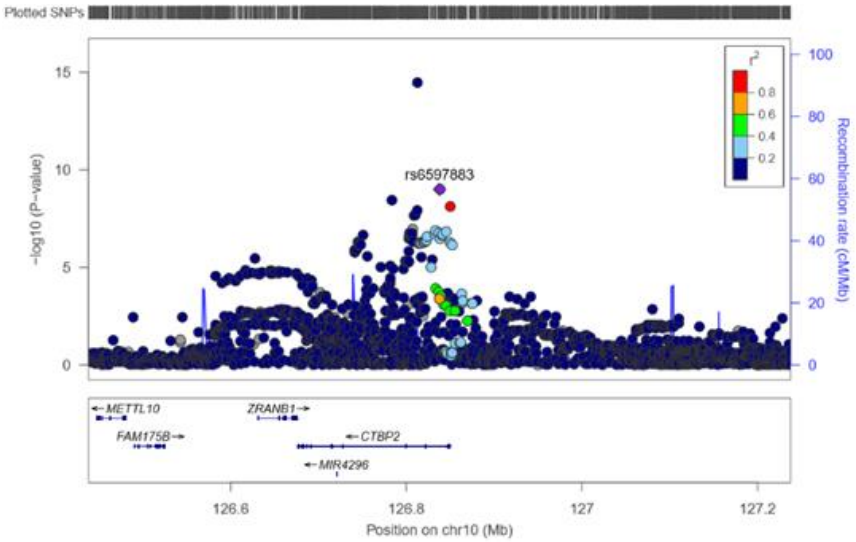

Figure t. Locus plot of genome region flanking SNP rs6597883

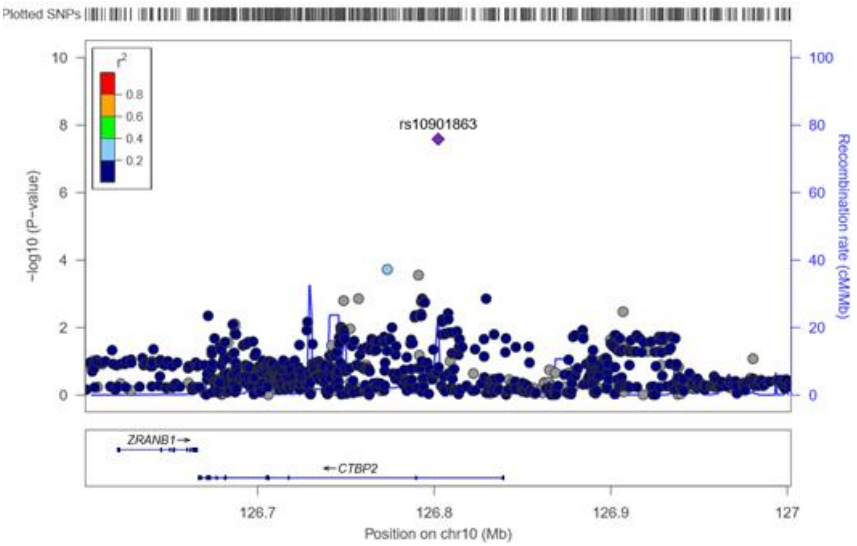

Figure u. Locus plot of genome region flanking SNP rs10901863

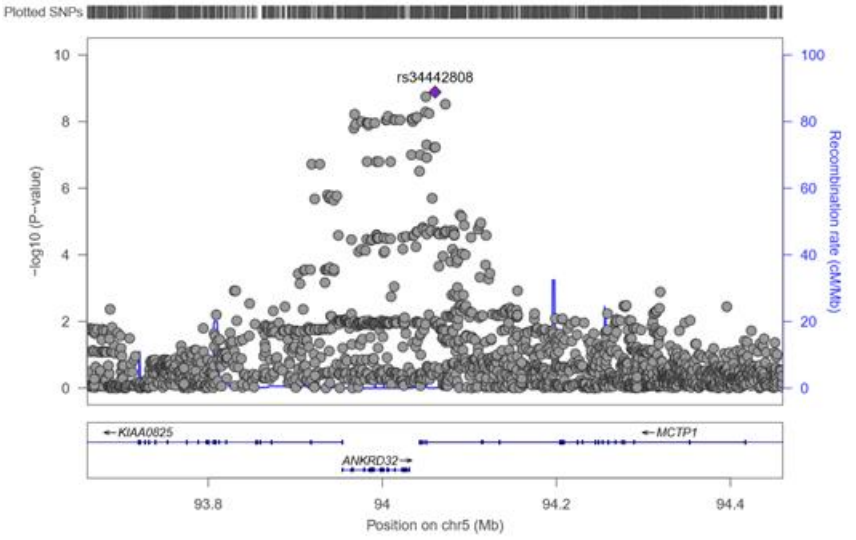

Figure v. Locus plot of genome region flanking SNP rs34442808

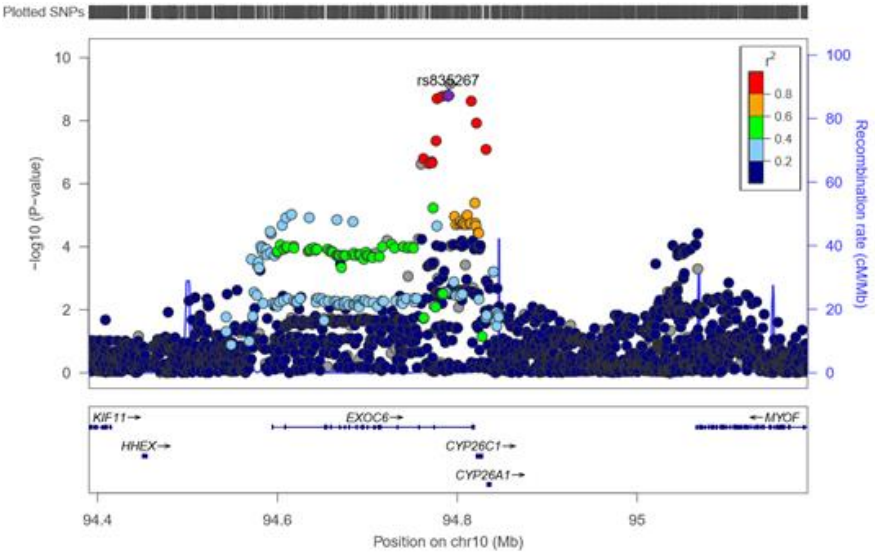

Figure w. Locus plot of genome region flanking SNP rs835267

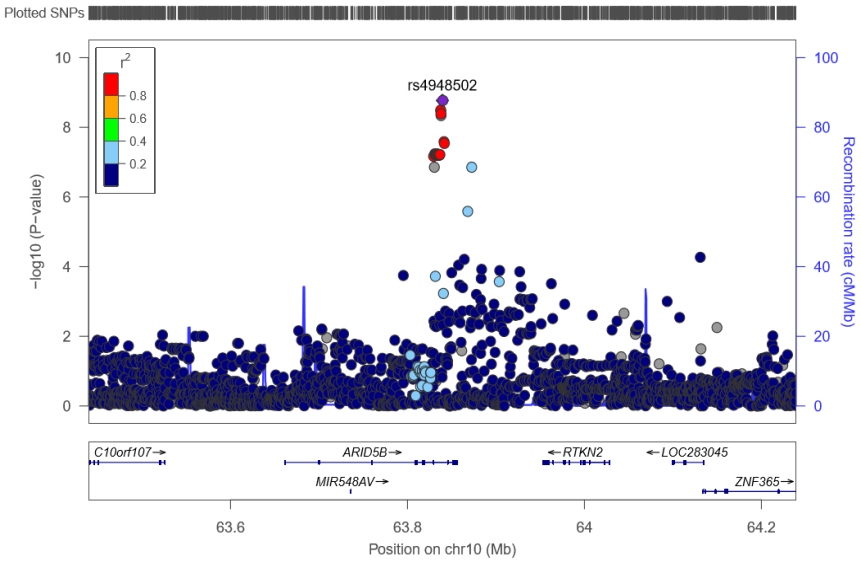

Figure x. Locus plot of genome region flanking SNP rs4948502

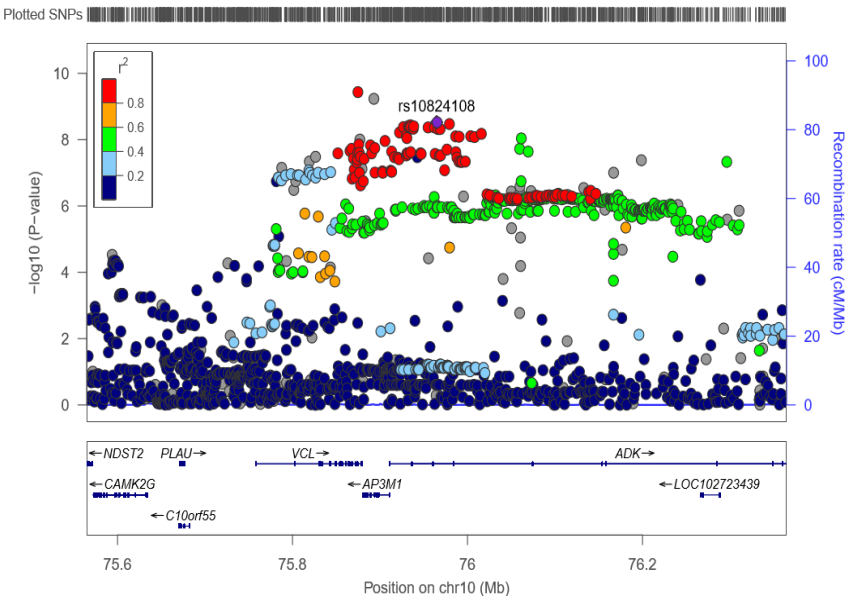

Figure y. Locus plot of genome region flanking SNP rs10824108

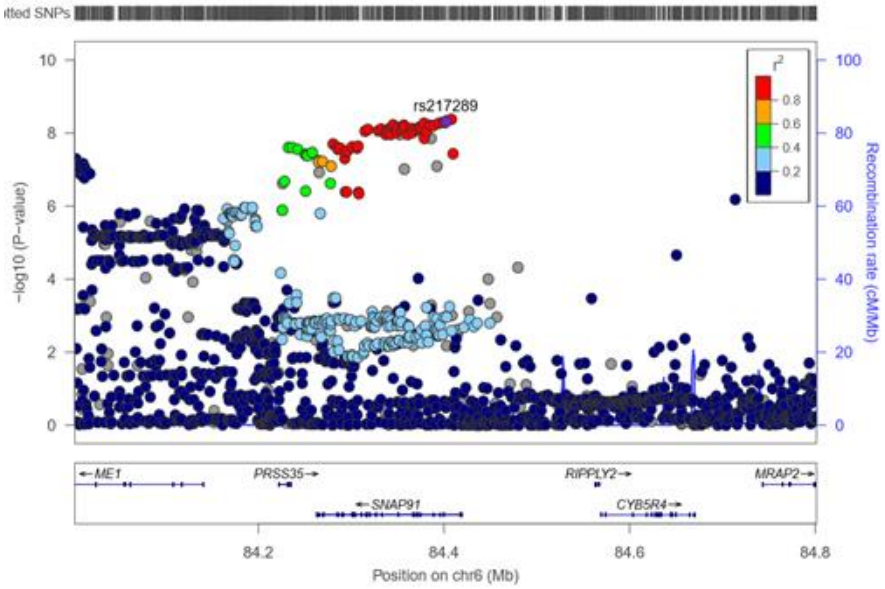

Figure z. Locus plot of genome region flanking SNP rs217289

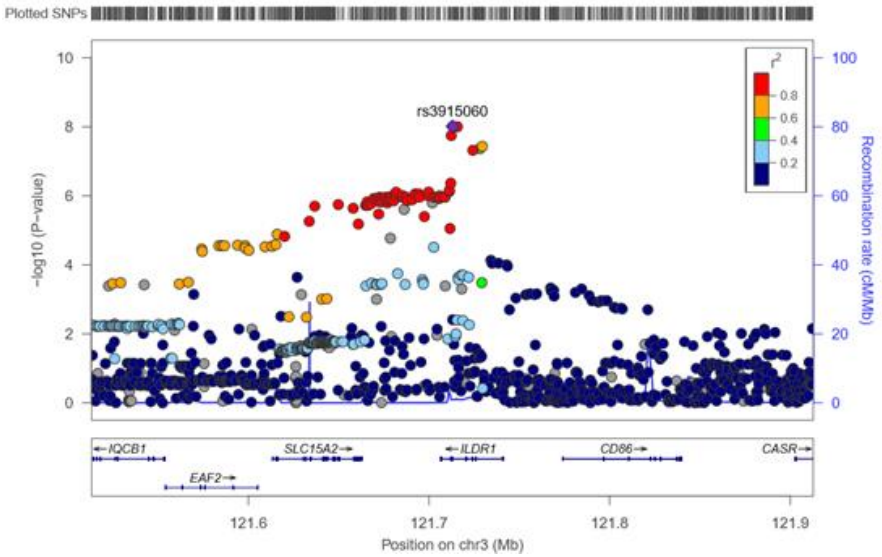

Figure aa. Locus plot of genome region flanking SNP rs3915060

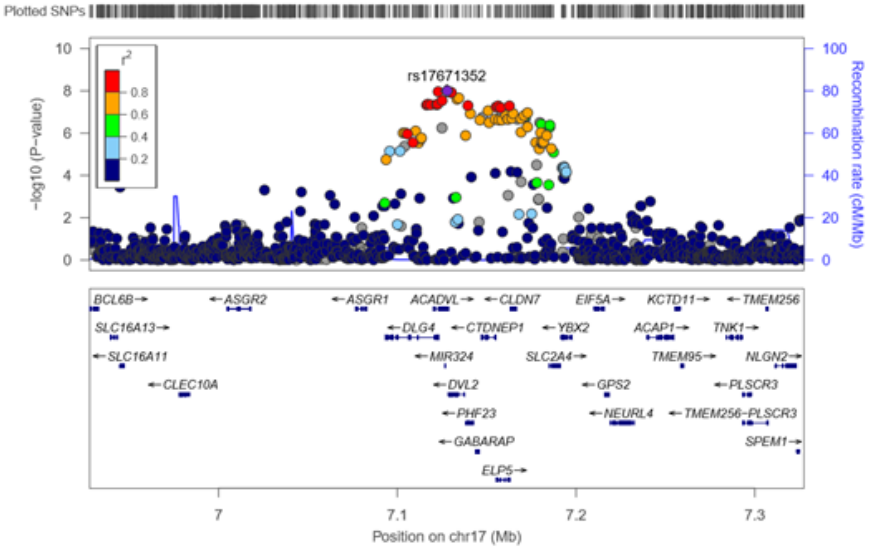

Figure bb. Locus plot of genome region flanking SNP rs17671352

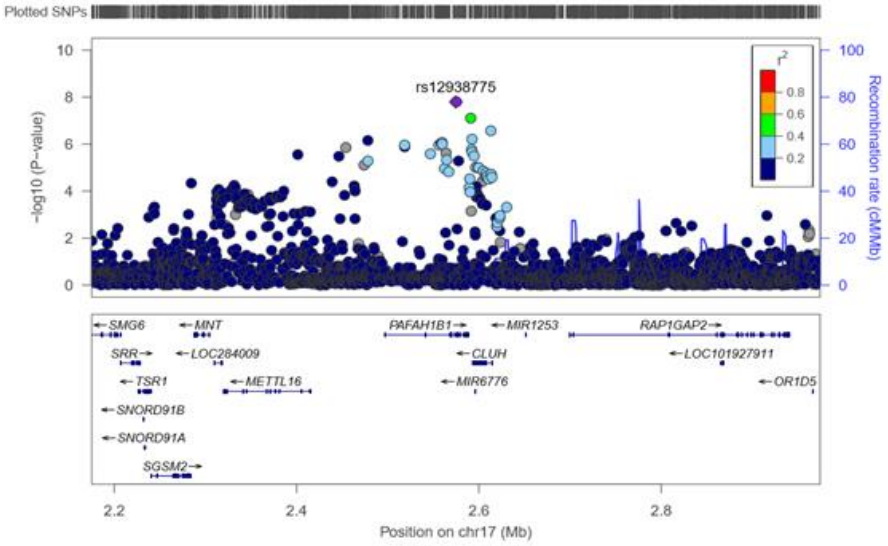

Figure cc. Locus plot of genome region flanking SNP rs12938775

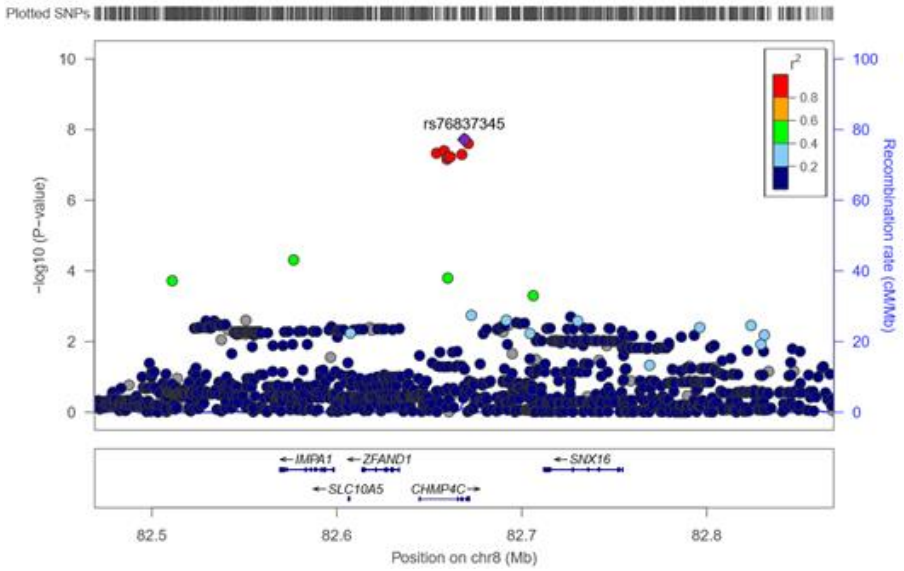

Figure dd. Locus plot of genome region flanking SNP rs76837345

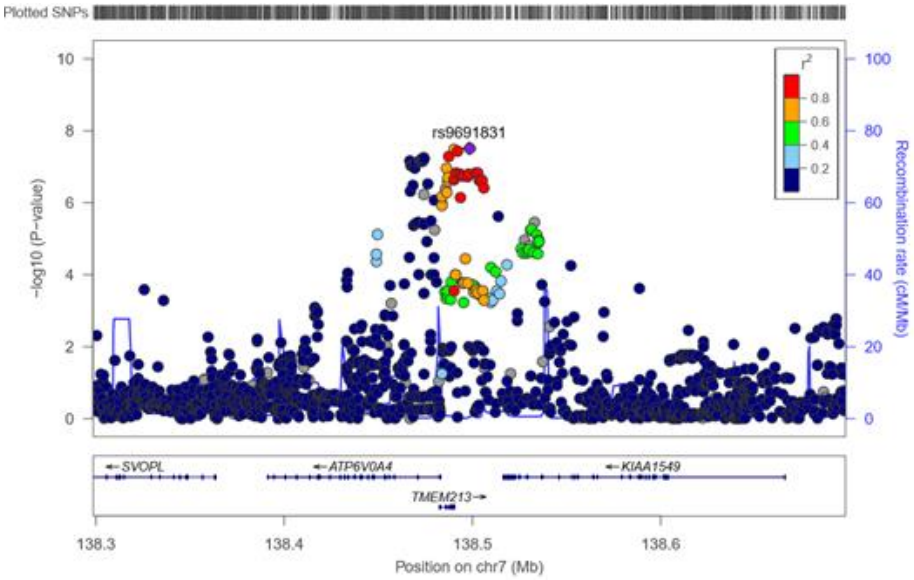

Figure ee. Locus plot for genome region flanking SNP rs9691831

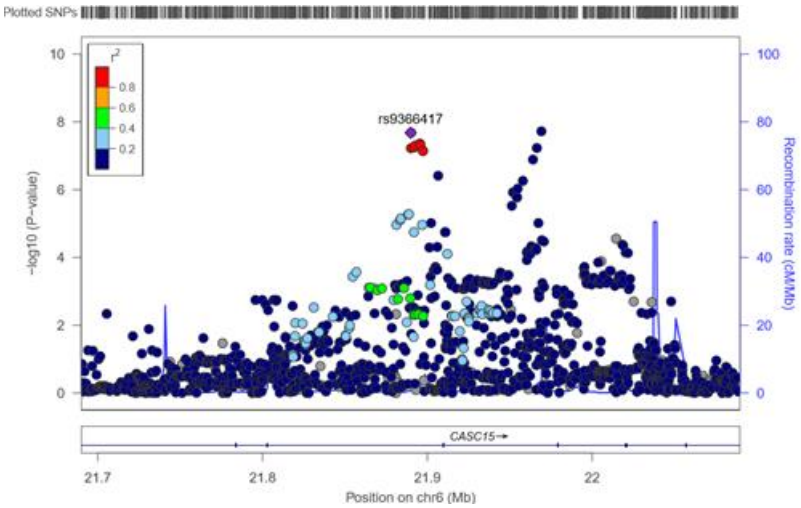

Figure ff. Locus plot for genome region flanking SNP rs9366417

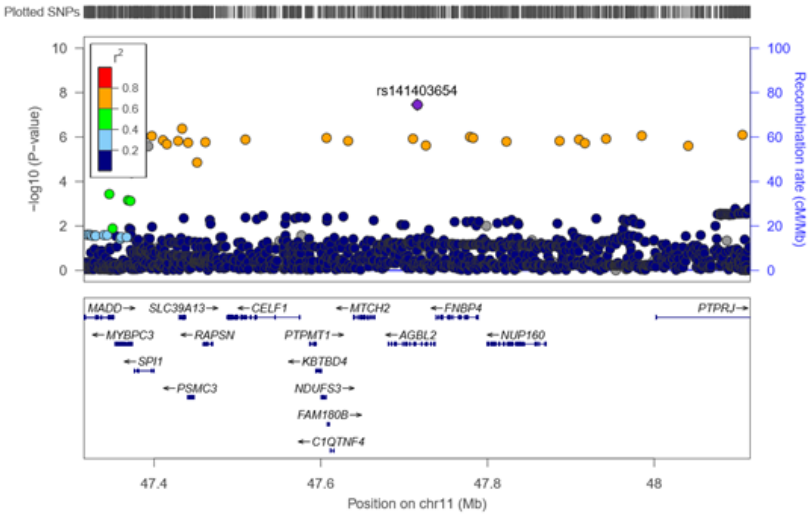

Figure gg. Locus plot for genome region flanking SNP rs141403654

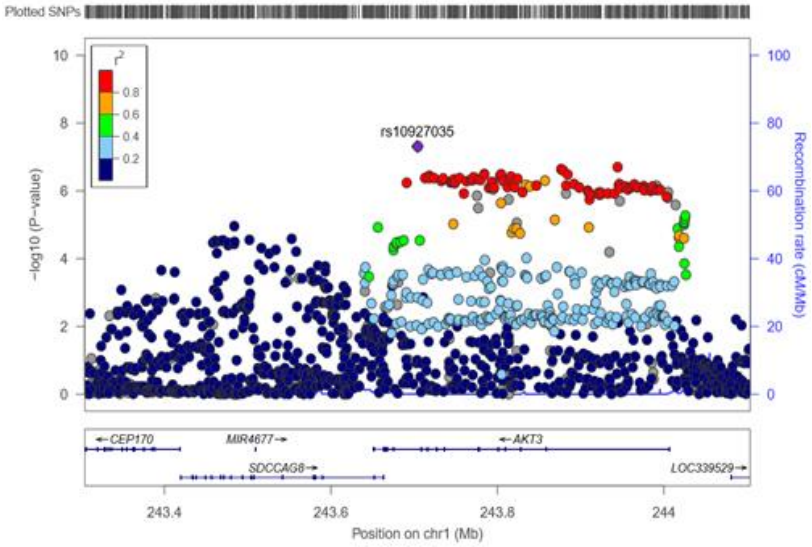

Figure hh. Locus plot for genome region flanking SNP rs10927035

Locus plots (ii-vv) for lead SNPs that reside in intergenic regions

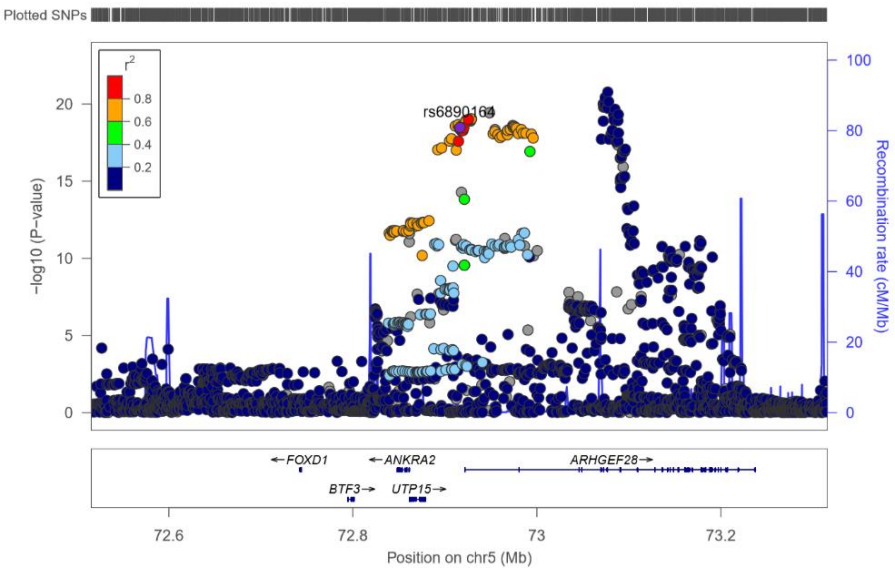

Figure ii. Locus plot for genome region flanking SNP rs6890164

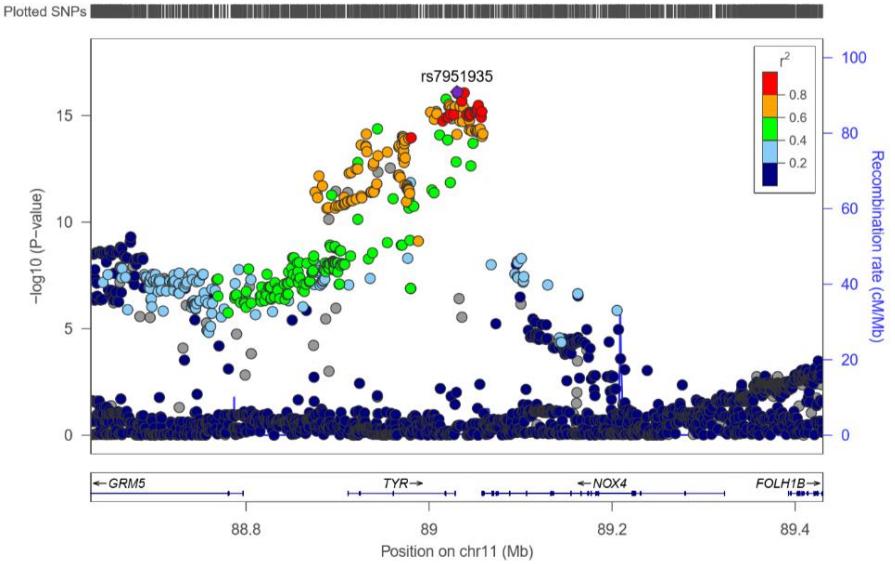

Figure jj. Locus plot for genome region flanking SNP rs7951935

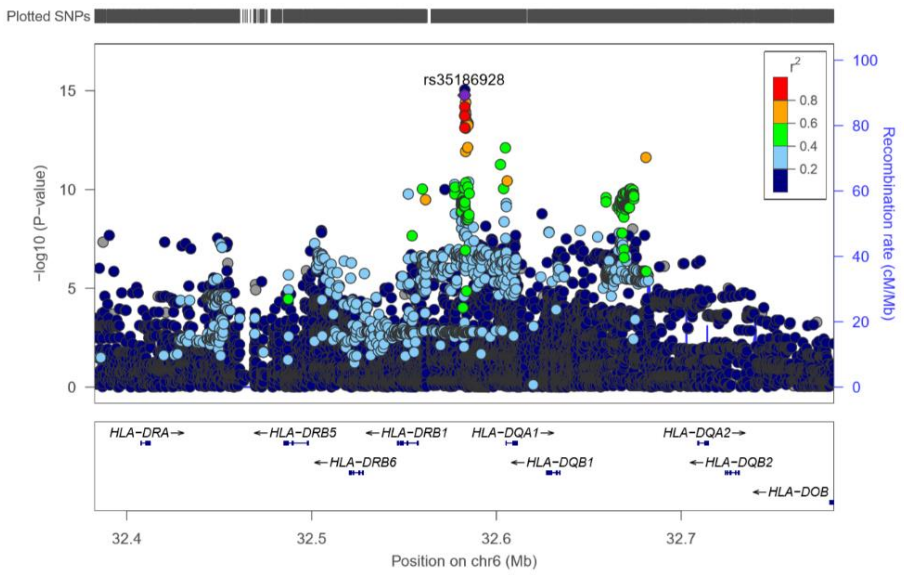

Figure kk. Locus plot for genome region flanking SNP rs35186928

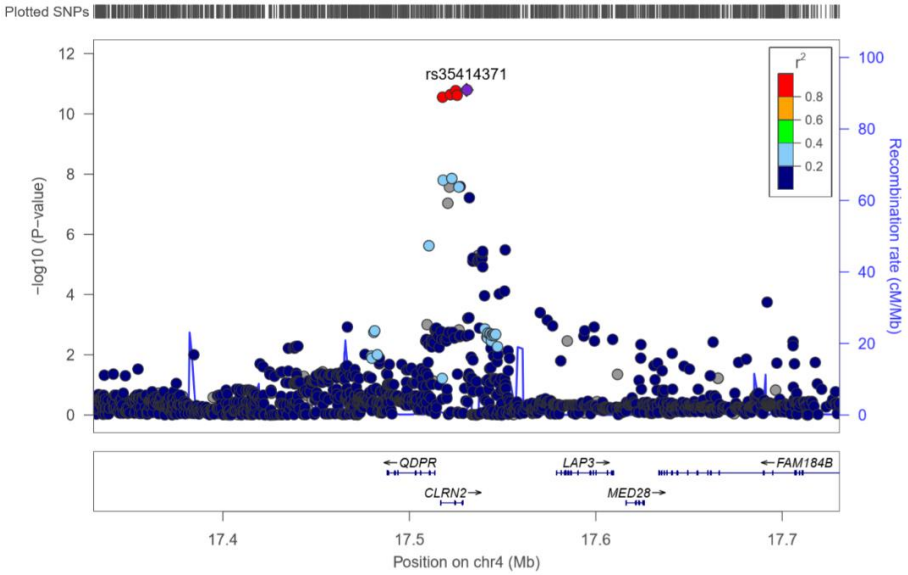

Figure ll. Locus plot for genome region flanking SNP rs3514371

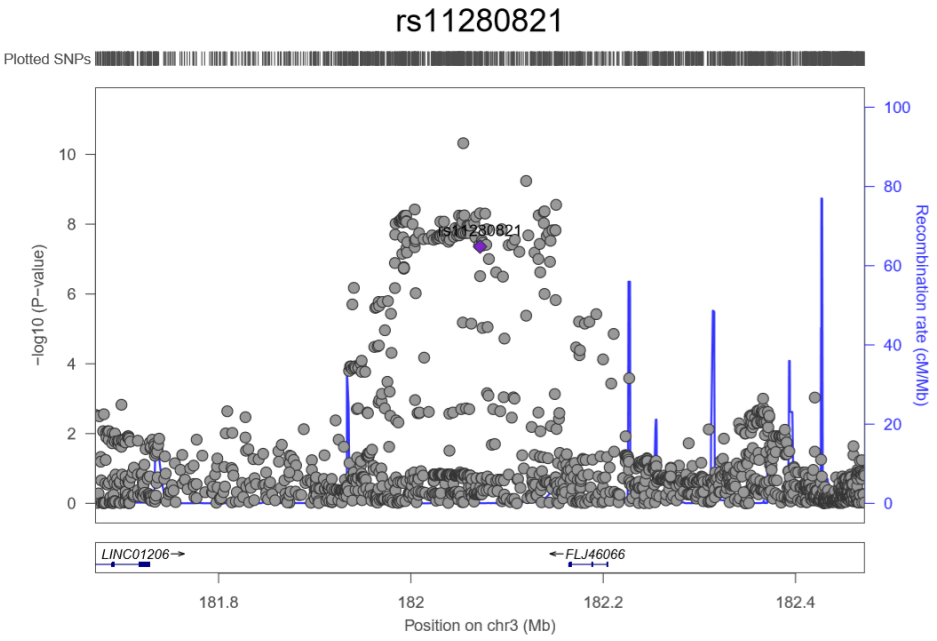

Figure mm. Locus plot for genome region flanking SNP rs11280821. Proximity to 3:182069497_TA_T

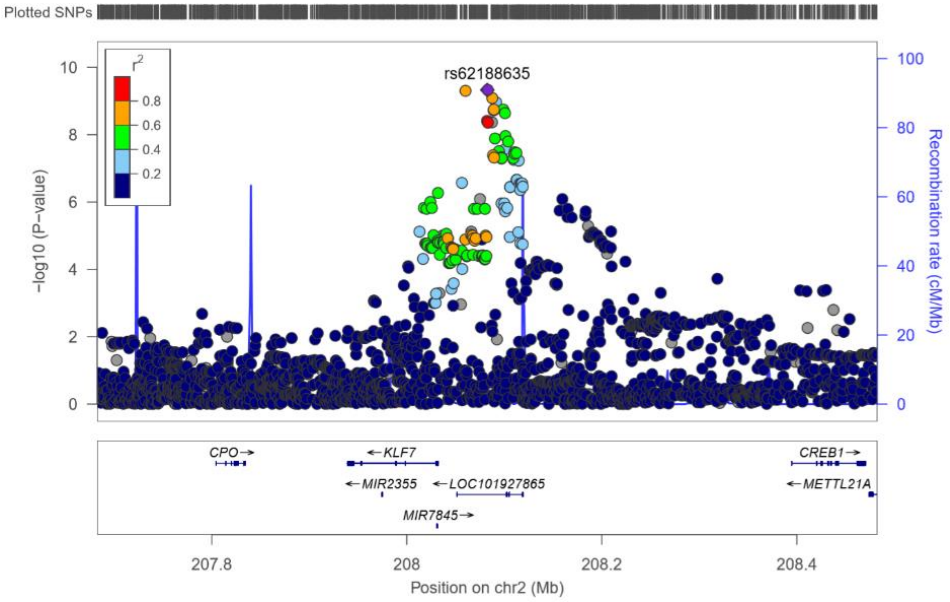

Figure nn. Locus plot for genome region flanking SNP rs62188635

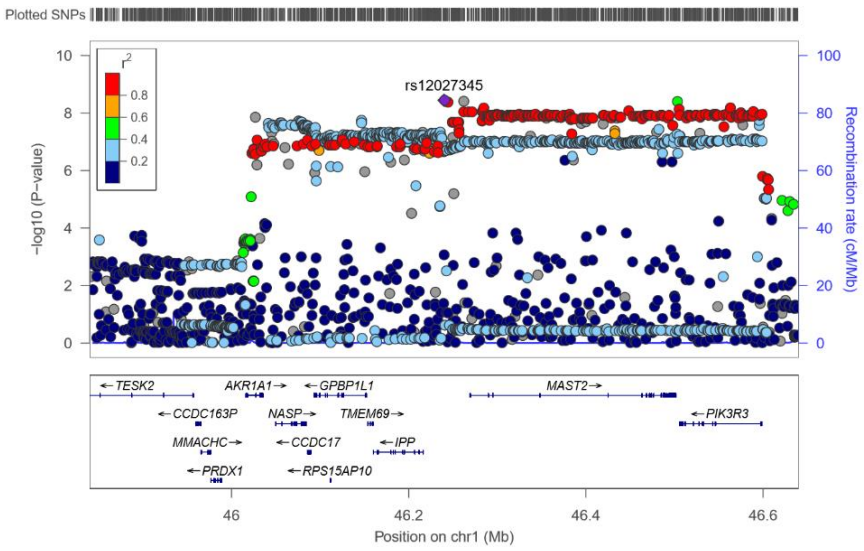

Figure oo. Locus plot for genome region flanking SNP rs12027345

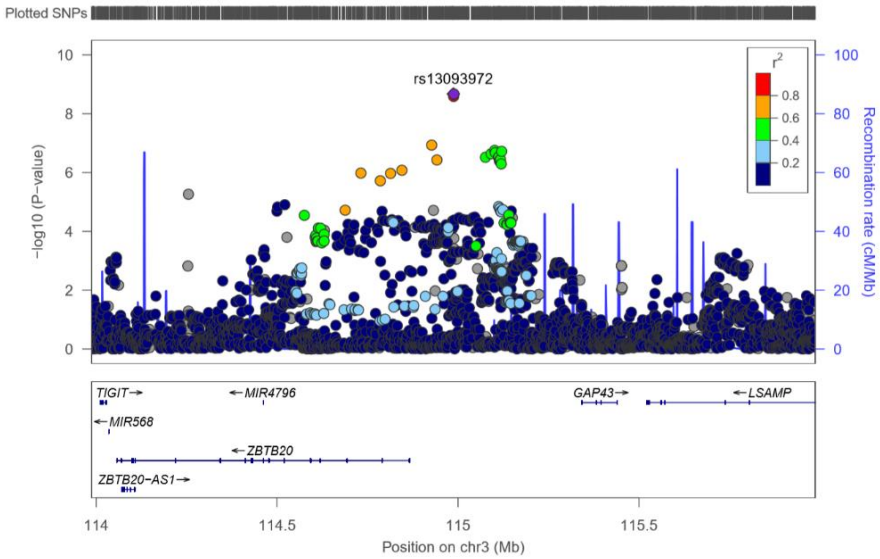

Figure pp. Locus plot of genome region flanking SNP rs13093972

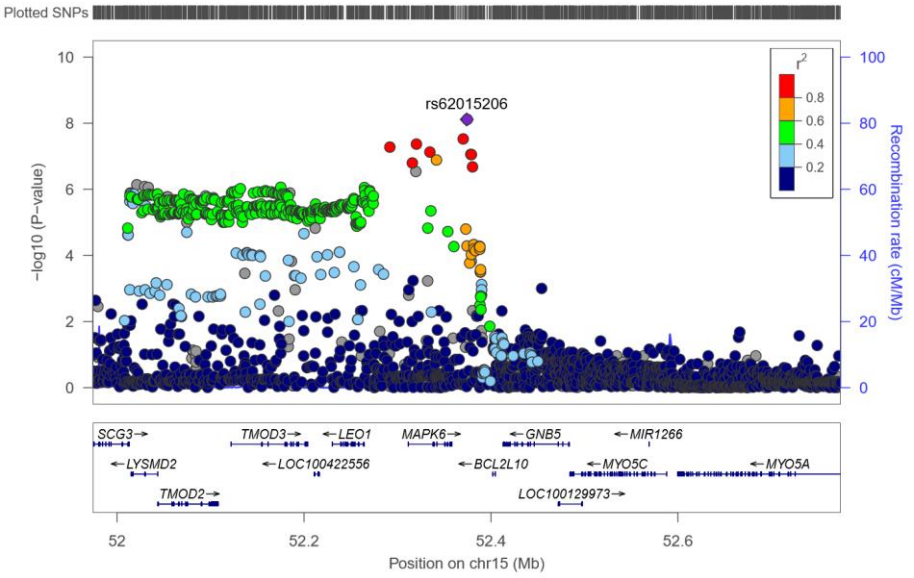

Figure qq. Locus plot of genome region flanking SNP rs62015206

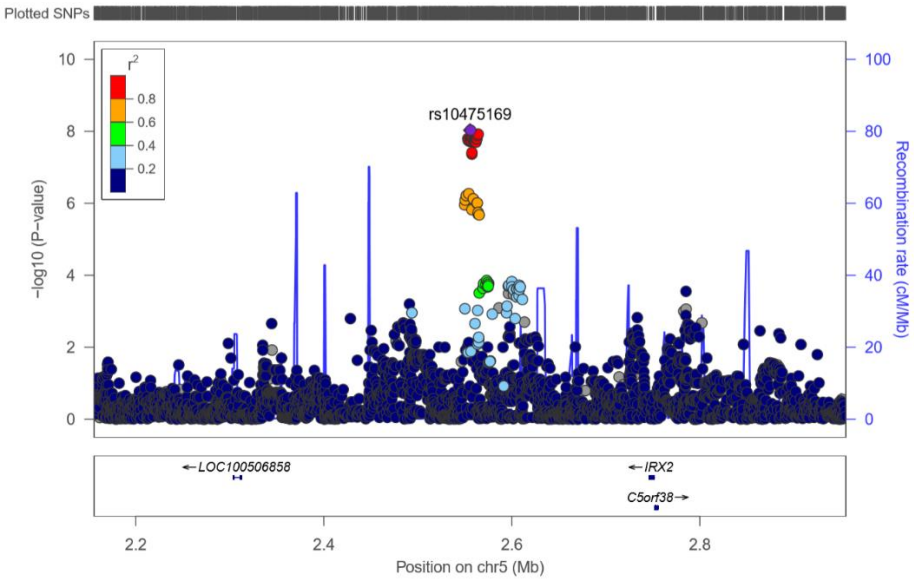

Figure rr. Locus zoom plot for genome region flanking SNP rs10475169

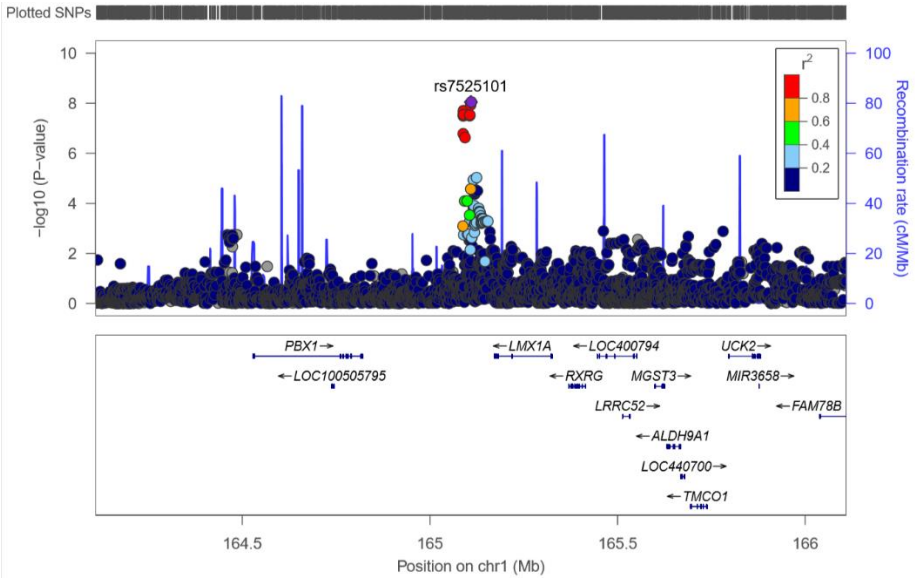

Figure ss. Locus zoom plot for genome region flanking rs7525101

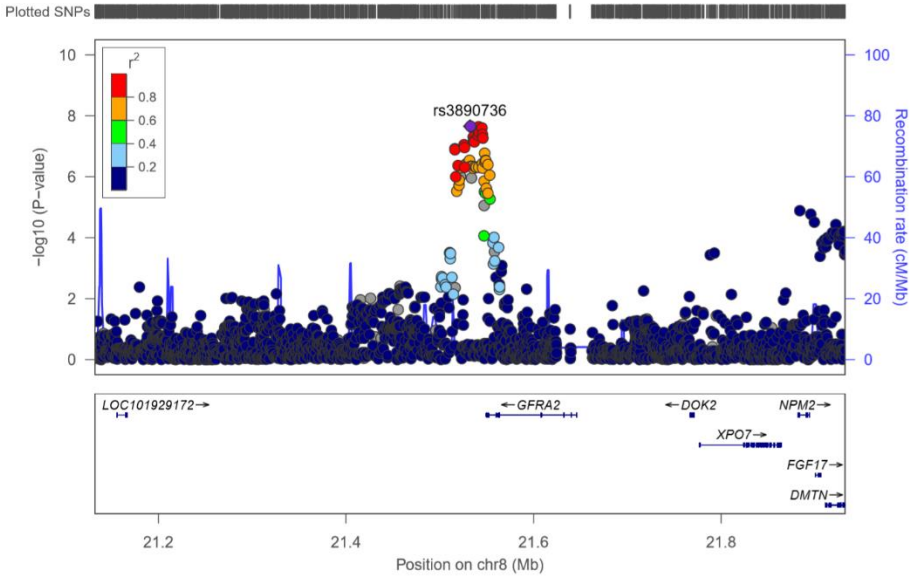

Figure tt. Locus zoom plot for genome region flanking SNP rs3890736

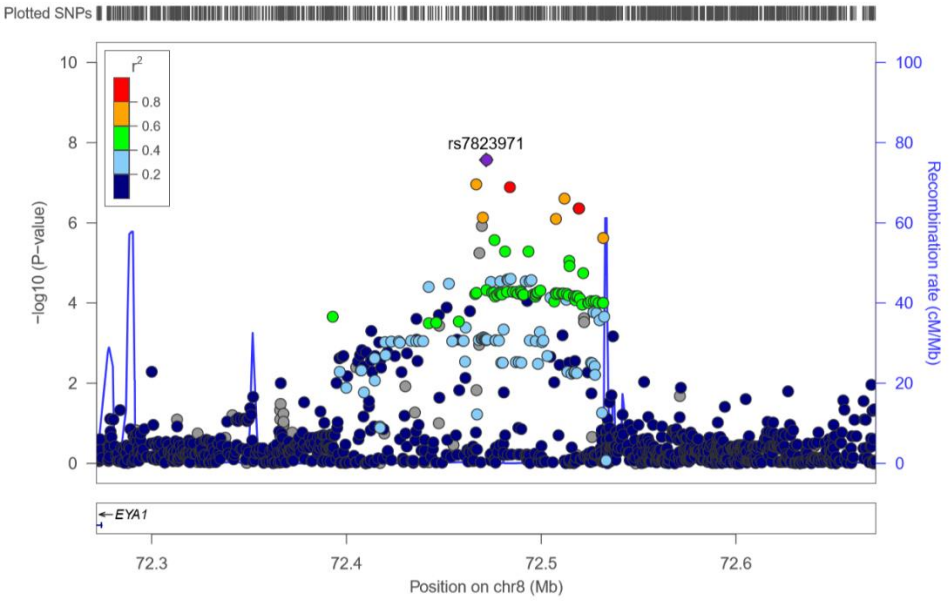

Figure uu. Locus plot for genome region flanking SNP rs7823971

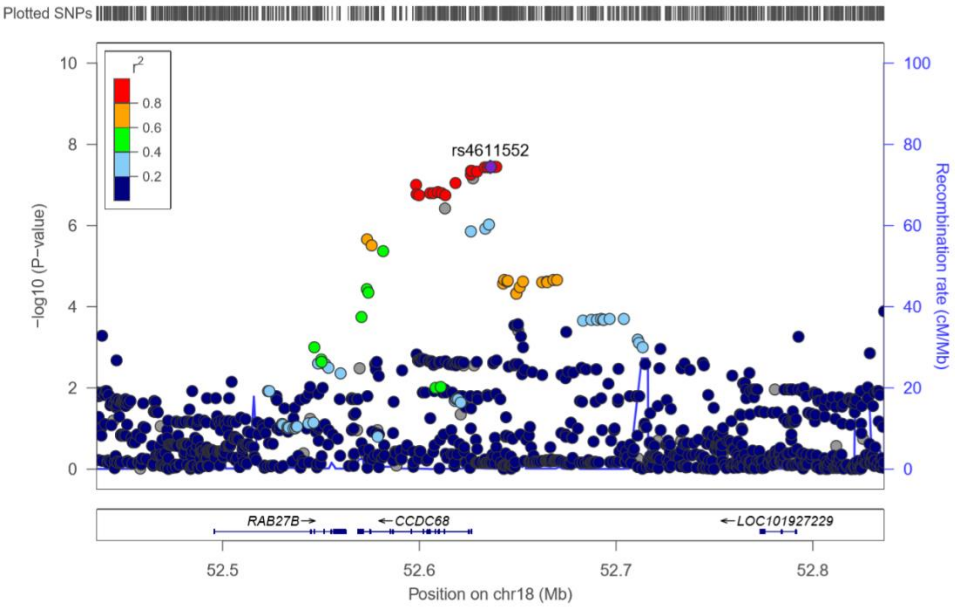

Figure vv. Locus plot of genome region flanking SNP rs4611552